AUTOCATALYSIS/YEAST TWO-HYBRID ASSAY

5   **FIELD OF THE INVENTION**

The present invention is related to an improved assay system wherein protein-protein interactions that require specific post-translational modifications, or are inhibited by specific post-translational modifications of the relevant proteins can be detected.

10   **BACKGROUND**

Protein-protein interactions are fundamental to proteomics (proteomics can be defined as the qualitative and quantitative studies of the proteome, the protein products of a species genome). Proteins are the most abundant and versatile macromolecules in living systems and serve crucial functions in essentially all biological processes.

15   Proteins may perform structural, transport, protective, catalytic, sensory, neuro-transmitting, regulatory and many other functions. Though versatile as they are, considering the complexity of even the simplest life form, it is not surprising that proteins rarely function by themselves. Rather, they interact with other proteins and molecules. In the proteomic and genomic era, it has become clear to researchers that

20   what constitutes a cell is the collective effort of these proteins whose functions are key to normal development and differentiation. Protein-protein interactions are fundamental to the understanding of biology, disease and even life itself.

In order to understand the normal biological processes and the diseases resulting from breakdown of normal functioning of proteins, it is important to study

25   protein interactions. In the last twenty years science has made significant progress in the study of protein-protein interactions. The techniques developed include protein precipitation, transfection of suspected interacting proteins into host cells, in vitro

1

biochemical analyses and yeast two-hybrid screening.

Thus far, the most versatile genetic system for screening and testing protein-protein interactions is the yeast two-hybrid (Y2H) system (Fields, S., and O. Song, "A novel genetic system to detect protein-protein interactions" *Nature* 340:245-246, 1989; U.S. Patent Nos. 5,283,273, 5,468,614 and 5,667,973). Currently, it is estimated that there are at least 10,000 interactions among the 6,000 proteins in yeast (Uetz, P., "Two-hybrid arrays" *Curr Opin Chem Biol* 6:57-62, 2002). The actual number of protein-protein interactions is probably much higher, because the two proteins, bait and prey, studied by the Y2H method are not designed to be modified (see below) in any way. Thus, protein-protein interactions that require either interacting protein to be chemically modified will escape the detection by the Y2H method.

Post-translational modifications, or PTMs, refer to the specific chemical moieties added to target amino acid residues of proteins after the latter are synthesized (translated). Numerous proteins contain specific PTMs that are critical for their functions. PTMs may activate or inactivate the recipient proteins. Certain PTMs may flag the modified proteins for degradation or transport to selective intra- or extra-cellular destiny. Many more PTMs perform yet to be identified functions. Common PTMs include acetylation, phosphorylation, methylation, ubiquitylation, glycosylation, etc. Frequently, these PTMs are indispensable for the functions of the recipient proteins. However, in most cases, it is not known exactly what these PTMs do at a molecular level. One well-thought idea is that specific PTMs create new interface for protein-protein interactions. Some of these interactions may occur only after one of the two interacting partners is modified at a particular amino acid residue(s). For example, the well-conserved PhosphoTyrosine Binding (PTB) motif interacts with proteins that are phosphorylated at various tyrosine residues. The bromodomain that is shared by many transcriptional activators binds histones that are acetylated (Dhalluin, C., *et al.*, "Structure and ligand of a histone acetyltransferase bromodomain" *Nature* 399:491-496, 1999; Jacobson, R. H., *et al.*, "Structure and function of a human TAFII250 double bromodomain module" *Science* 288:1422-1425,

2000). In contrast, it is equally possible that existing protein-protein interactions may be inhibited if one of the two interacting proteins receives a particular modification. For example, the Silent Information Regulator protein Sir3 binds only to unacetylated histones for transcriptional repression (Edmondson, D. G., M. M. Smith, and S. Y. Roth, "Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4" *Genes Dev* 10:1247-1259). Acetylation of the histones antagonizes the function of Sir3 and leads to transcriptional de-silencing of the underlying genes (Carmen, *et al.*, "Acetylation of the yeast histone H4 N terminus regulates its binding to heterochromatin protein SIR3" *J Biol Chem* 277:4778-81, 2002). These biochemical data support the idea that PTMs may positively or negatively regulate protein-protein interactions. However, these reports only represent sporadic examples of such regulation. In other words, to understand how PTMs regulate protein-protein interactions at a global and proteomic scale, a non-biased genetic method is needed.

In light of the primal importance of the possible effects of PTMs on protein-protein interactions, several groups independently reported a common strategy with which these authors were able to detect protein-protein interactions induced by specific phosphorylation (Cao, H., W. E. Courchesne, and C. C. Mastick, "A phosphotyrosine-dependent protein interaction screen reveals a role for phosphorylation of caveolin-1 on tyrosine 14: recruitment of C-terminal Src kinase" *J Biol Chem* 277:8771-8774, 2002; Shaywitz, A. J., S. L. Dove, M. E. Greenberg, and A. Hochschild, "Analysis of phosphorylation-dependent protein-protein interactions using a bacterial two-hybrid system" *Sci STKE* 2002:L11, 2002; Yamada, M., *et al.*, "Analysis of tyrosine phosphorylation-dependent protein-protein interactions in TrkB-mediated intracellular signaling using modified yeast two-hybrid system" *J Biochem (Tokyo)* 130:157-65, 2001). In each case, kinases were expressed in the two-hybrid reporter cells that normally lack such enzyme systems. The bait proteins produced in these cells are thus modified by the foreign enzymes. For example, a tyrosine kinase and a serine/threonine kinase were expressed in yeast Shaywitz, A. J., S. L. Dove, M. E. Greenberg, and A. Hochschild, "Analysis of phosphorylation-dependent protein-

3

protein interactions using a bacterial two-hybrid system" *Sci STKE* 2002:L11, 2002;

Yamada, M., *et al.*, "Analysis of tyrosine phosphorylation-dependent protein-protein interactions in TrkB-mediated intracellular signaling using modified yeast two-hybrid system" *J Biochem (Tokyo)* 130:157-65, 2001) and E. coli (Cao, H., W. E.

Courchesne, and C. C. Mastick, "A phosphotyrosine-dependent protein interaction screen reveals a role for phosphorylation of caveolin-1 on tyrosine 14: recruitment of C-terminal Src kinase" *J Biol Chem* 277:8771-8774, 2002), respectively. The substrate proteins, existing in the form of two-hybrid baits, were shown to be phosphorylated in vivo and consequently allowed the detection of interactions involving protein phosphorylation Yamada, M., *et al.*, "Analysis of tyrosine phosphorylation-dependent protein-protein interactions in TrkB-mediated intracellular signaling using modified yeast two-hybrid system" *J Biochem (Tokyo)* 130:157-65, 2001).

The above methods rely on the typical enzyme-substrate reactions occurring in trans (*i.e.*, between two distinct proteins) to create the baits for genetic selection. Therefore, one concern is whether the efficiency of the bait modification would be sufficient to surpass the level of the unmodified bait, permitting a high signal-to-noise ratio in the genetic screen. To avoid this potential problem, one may choose to over-produce the foreign enzyme. However, such treatment may result in uncontrolled enzymatic action on host proteins, leading to cellular toxicity.

Unfortunately, there is not any method that enables researchers to screen for such interactions in a global, non-biased manner. Therefore, what is needed is an assay system that enables researchers to detect protein-protein interactions that are dependent upon or inhibited by post-translational modifications.

SUMMARY OF THE INVENTION

The AC/Y2H system of the present invention offers two major advantages over present technologies. In one embodiment the enzyme is expressed in its natural host where the opposing enzymes (*i.e.*, HDACs) are present. Pleiotropic effect is less likely. On the other hand, the desired catalysis is most likely carried out in *cis*, and is dominant over the endogenous HDAC trans-activity, leading to a constitutive

4

modification of the bait. Moreover, the use of the catalytically inactivated enzyme in a parallel, counter screen will allow fast identification of protein interactions disrupted by selective PTMs. A reversal in the genetic screening criteria can reveal protein interactions that are perturbed by a specific PTM.

5   A simple modification of the current AC/Y2H constructs may identify other proteins that recognize different histone modifications or different modifications of other proteins. For example, in one embodiment, a substitution of the HAT with other histone modifying enzymes such as, for example, kinases (see, for example, Table 2) will create AC baits containing the corresponding modifications. AC/Y2H can thus be
10   used to identify the cognate binding factors. It this way, it is contemplated that the present invention can detect the PTMs of other proteins that may be subjected to similar studies.

In one embodiment, the present invention contemplates a compound, comprising a) a first amino acid sequence comprising at least a portion of a histone
15   amino terminal tail, said first amino acid sequence linked to b) a second amino acid sequence comprising at least a portion of a histone acetyltransferase. In another embodiment, the present invention contemplates that the second amino acid sequence comprises the active catalytic domain of Gcn5 (see, for example, Figure 20, [SEQ ID NO: 23] from plasmid pDG28 [SEQ ID NO: 9] and Figure 22, [SEQ ID NO: 25] from
20   plasmid pDG30 [SEQ ID NO: 11]). In yet another embodiment, the present invention contemplates that the second amino acid sequence comprises a catalytically inactive portion of Gcn5 (see, for example, Figure 21, [SEQ ID NO: 24] from plasmid pDG29 [SEQ ID NO:10] and Figure 23, [SEQ ID NO: 26] from plasmid pDG31 [SEQ ID NO: 12]). In still yet another embodiment, the present invention contemplates that the
25   first amino acid sequence comprises the histone H3 tail (see, for example, Figure 12, pDG1 [SEQ ID NO: 15]; Figure 13, pDG2 [SEQ ID NO: 16]; Figure 16, pDG5 [SEQ ID NO: 19] and; Figure 17, pDG6 [SEQ ID NO: 20]). In still yet another embodiment, the present invention contemplates that the first amino acid sequence comprises the histone H4 tail (amino acids 1 - 29) (see, for example, Figure 14, pDG3
30   [SEQ ID NO: 17]; Figure 15, pDG4 [SEQ ID NO: 18]; Figure 18, pDG7 [SEQ ID

NO: 21] and; Figure 19, pDG8 [SEQ ID NO: 22]). In still yet another embodiment, the present invention contemplates that the compound comprises a fusion protein. In still yet another embodiment, the present invention contemplates that the compound exhibits autoacetylation. In still yet another embodiment, the present invention contemplates that the compound further comprises a DNA binding moiety. In still yet another embodiment, the present invention contemplates that the DNA binding moiety is linked to said first amino acid sequence. In still yet another embodiment, the present invention contemplates that the DNA binding moiety comprises the Gal4 DNA binding domain. In still yet another embodiment, the present invention contemplates that the compound further comprises a detectable moiety linked to said second amino acid sequence. In still yet another embodiment, the present invention contemplates that the detectable moiety comprises an epitope.

In one embodiment, the present invention contemplates a nucleic acid encoding the fusion protein of the present invention. In another embodiment, the present invention contemplates an expression vector comprising the nucleic acid. In yet another embodiment, the present invention contemplates yeast transformed with the expression vector. In still yet another embodiment, the present invention contemplates a whole cell extract of the yeast.

In one embodiment, the present invention contemplates a method for detecting protein-protein interactions, said interactions requiring a post translational modification of one of the said proteins, said method comprising: (a) providing a host cell comprising a detectable gene wherein the detectable gene expresses a detectable protein when the detectable gene is activated by an amino acid sequence including a transcriptional activation domain when the transcriptional activation domain is in sufficient proximity to the detectable gene; (b) providing a first chimeric gene that is capable of being expressed in the host cell, the first chimeric gene comprising a DNA sequence that encodes a first hybrid protein, the first hybrid protein comprising: (i) a DNA-binding moiety that recognizes a binding site on the detectable gene in the host cell; (ii) a first test protein or fragment thereof, comprising a reactive moiety capable of being modified through catalysis, that is to be tested for interaction with at least one

6

second test protein or fragment thereof; and (iii) a catalytic moiety that is capable of catalyzing said first test protein; (c) providing a second chimeric gene that is capable of being expressed in the host cell, the second chimeric gene comprising a DNA sequence that encodes a second hybrid protein, the second hybrid protein comprising:

5   (i) the transcriptional activation domain; and (ii) a second test protein or fragment thereof that is to be tested for interaction with the first test protein or fragment thereof when said first test protein has been modified by the catalysis of said reactive moiety to create a modified first test protein; wherein interaction between the first modified test protein and the second test protein in the host cell causes the transcriptional

10   activation domain to activate transcription of the detectable gene; (d) introducing the first chimeric gene and the second chimeric gene into the host cell; (e) subjecting the host cell to conditions under which the first hybrid protein and the second hybrid protein are expressed in sufficient quantity for the detectable gene to be activated; and (f) determining whether the detectable gene has been expressed to a degree greater

15   than expression in the absence of an interaction between the first test protein and the second test protein.

In another embodiment, the present invention contemplates a method, wherein said binding DNA-moiety comprises GDBD; said catalytic moiety comprises the catalytic domain of Gcn5 and; said reactive moiety comprises a histone amino terminal

20   tail capable of being acetylated by Gcn5. In yet another embodiment, the present invention contemplates a method, wherein said first test protein and said second test protein are encoded on a library of plasmids containing DNA inserts, derived from the group consisting of genomic DNA, cDNA and synthetically generated DNA. In still yet another embodiment, the present invention contemplates a method, wherein first

25   test protein and said second test protein are derived from derived from the group consisting of bacterial protein, viral protein, oncogene-encoded protein, mammalian protein, fungal protein and plant protein.

In one embodiment, the present invention contemplates that the a compound, comprising a) a first amino acid sequence comprising at least a portion of a histone

30   amino terminal tail, said first amino acid sequence linked to b) a second amino acid

sequence comprising at least a portion of a histone modifying enzyme. In another embodiment, the present invention contemplates a compound, wherein said histone modifying enzyme comprises at least a portion of a acetyltransferase. In yet another embodiment, the present invention contemplates a compound, wherein said portion

5   comprises an active catalytic domain of an acetyltransferase.

One embodiment of the design of the AC/2H system is illustrated in Figure 1. The present invention is not limited to this embodiment. Other embodiments are also contemplated and contained herein. Other the two-hybrid interactions and their dependence on the modification of the protein of interest (A) are in Table 1. This

10   embodiment of the system is composed of the following elements:

In one embodiment, the protein A that is known or suspected to be the substrate of protein B (e.g., an enzyme) for a post-translational modification is physically fused to the enzyme B. This fusion is achieved by ligating in-frame the DNA fragments encoding A and B. The A-B fusion is expressed in bacterial or yeast

15   cells (alone or within the context of two-hybrid baits, see below), the protein B is able to affect (e.g., catalyze) modification of the linked protein A resulting in the "autocatalysis" of the A-B hybrid protein.

In one embodiment, the A-B DNA fusion is further ligated in-frame to another DNA fragment which encodes a protein module C that, e.g., constitutes the bait or the

20   prey hybrid protein in the two-hybrid system. For example, in the Yeast Two-Hybrid system (U.S. Patent Nos 5,283,273, 5,468,614 and 5,667,973), the module would be, for example, the DNA binding domain of Gal4 or LexA, or the activation domain of a transcriptional activator, whereas in the Spit-Ubiquitin system (U.S. Patent Nos. 5,503,977 and 5,585,245), module C would be, for example, the NubG or Cub

25   (Johnsson, et al., "Split ubiquitin as a sensor of protein interactions in vivo" Proc. Natl. Acad. Sci. USA 91:10340-10344, 1994).

In one embodiment, the C-A-B hybrid DNA fragment is further ligated in-frame to module D which encodes a peptide used as an epitope tag that can be detected and quantified by immunochemical means.

30   In one embodiment, an otherwise identical C-A-B'-D hybrid DNA is created in

the same manner as the C-A-B-D hybrid. The B' fragment encodes a catalytically
inactive form of the enzyme B which is created by site-directed mutagenesis to ablate
the catalytic power of the enzyme B. Therefore, while the C-A-B-D chimera contains
a constitutive modification within the A module, the C-A-B'-D mutant enzyme fusion
fails to do so due to the mutation(s) (*e.g.*, point mutation(s)) that abolishes the
catalytic power of the enzyme B.

In one embodiment, the C-A-B-D and C-A-B'-D chimeric DNA fragments are
each ligated to a plasmid vector designed for the corresponding two-hybrid system.

In one embodiment, in addition to the C-A-B-D and C-A-B'-D chimera, two
additional control hybrids are created: C-B-D and C-B'-D.

In one embodiment, the final two-hybrid plasmids bearing the in-frame fusion
of C-A-B-D, C-A-B'-D, C-B-D, or C-B'-D hybrid are delivered (transformed) to the
corresponding two-hybrid host cells, and the quantity of the four hybrids and the
modification status of protein A in both cases are characterized by appropriate means,
such as immunochemical analyses using antibodies specific for D, and antibodies
specific for A that is modified by the enzyme B.

Although the present invention is not limited to any particular protein-protein
interaction, as summarized in Table 1, protein-protein interactions that are detected by
the set of autocatalysis baits can be one of four classes: Positive interactions with only
the C-A-B-D chimera, but not C-A-B'-D, C-B-D, or C-B'-D, are triggered by the
modification of A.; interactions that are detected by the C-A-B'-D hybrid, but not
C-A-B-D, C-B-D, or C-B'-D are specific for the unmodified A (*i.e.*, inhibited by A
modification by the enzyme B). Therefore, the AC/2H method is capable of detecting
protein-protein interactions that are either induced or inhibited by a specific
post-translational modification.

In one embodiment, the present invention contemplates a compound,
comprising a) a first amino acid sequence comprising at least a portion of a histone
amino terminal tail, said first amino acid sequence linked to b) a second amino acid
sequence comprising at least a portion of a histone acetyltransferase. In another
embodiment, the present invention contemplates that the second amino acid sequence

9

comprises the active catalytic domain of Gcn5. In yet another embodiment, the present invention contemplates that the second amino acid sequence comprises a catalytically inactive portion of Gcn5. In still yet another embodiment, the present invention contemplates that the first amino acid sequence comprises the histone H3 tail. In still yet another embodiment, the present invention contemplates that the first amino acid sequence comprises the histone H4 tail. In still yet another embodiment, the present invention contemplates that the compound comprises a fusion protein. In still yet another embodiment, the present invention contemplates that the compound exhibits autoacetylation. In still yet another embodiment, the present invention contemplates that the compound further comprises a DNA binding moiety. In still yet another embodiment, the present invention contemplates that the compound comprising the DNA binding moiety is linked to said first amino acid sequence. In still yet another embodiment, the present invention contemplates that the DNA binding moiety comprises the Gal4 DNA binding domain. In still yet another embodiment, the present invention contemplates that the DNA binding moiety comprises the Gal4 DNA binding domain further comprises a detectable moiety linked to said second amino acid sequence. In still yet another embodiment, the present invention contemplates that the detectable moiety comprises an epitope. In still yet another embodiment, the present invention contemplates the nucleic acid encoding the compounds of the present invention. In still yet another embodiment, the present invention contemplates an expression vector encoding the compounds of the present invention. In still yet another embodiment, the present invention contemplates yeast transformed with the expression vector. In still yet another embodiment, the present invention contemplates a whole cell extract of the yeast transfected with the expression vectors of the present invention.

In one embodiment, the present invention contemplates a method for detecting protein-protein interactions, said interactions requiring a post translational modification of one of the said proteins, said method comprising: (a) providing a host cell comprising a detectable gene wherein the detectable gene expresses a detectable protein when the detectable gene is activated by an amino acid sequence including a

10

transcriptional activation domain when the transcriptional activation domain is in sufficient proximity to the detectable gene; (b) providing a first chimeric gene that is capable of being expressed in the host cell, the first chimeric gene comprising a DNA sequence that encodes a first hybrid protein, the first hybrid protein comprising: (i) a DNA-binding moiety that recognizes a binding site on the detectable gene in the host cell; (ii) a first test protein or fragment thereof, comprising a reactive moiety capable of being modified through catalysis, that is to be tested for interaction with at least one second test protein or fragment thereof; and (iii) a catalytic moiety that is capable of catalyzing said first test protein; (c) providing a second chimeric gene that is capable of being expressed in the host cell, the second chimeric gene comprising a DNA sequence that encodes a second hybrid protein, the second hybrid protein comprising: (i) the transcriptional activation domain; and (ii) a second test protein or fragment thereof that is to be tested for interaction with the first test protein or fragment thereof when said first test protein has been modified by the catalysis of said reactive moiety to create a modified first test protein; wherein interaction between the first modified test protein and the second test protein in the host cell causes the transcriptional activation domain to activate transcription of the detectable gene; (d) introducing the first chimeric gene and the second chimeric gene into the host cell; (e) subjecting the host cell to conditions under which the first hybrid protein and the second hybrid protein are expressed in sufficient quantity for the detectable gene to be activated; and (f) determining whether the detectable gene has been expressed to a degree greater than expression in the absence of an interaction between the first test protein and the second test protein.

In another embodiment, the present invention contemplates that the binding DNA-moiety comprises GDBD; said catalytic moiety comprises the catalytic domain of Gcn5 and; said reactive moiety comprises a histone amino terminal tail capable of being acetylated by Gcn5. In yet another embodiment, the present invention contemplates that the first test protein and said second test protein are encoded on a library of plasmids containing DNA inserts, derived from the group consisting of genomic DNA, cDNA and synthetically generated DNA. In still yet another

embodiment, the present invention contemplates that the first test protein and said second test protein are derived from derived from the group consisting of bacterial protein, viral protein, oncogene-encoded protein, mammalian protein, fungal protein and plant protein.

In one embodiment, the present invention contemplates a compound, comprising a) a first amino acid sequence comprising at least a portion of an enzyme substrate, said first amino acid sequence linked to b) a second amino acid sequence comprising at least a portion of an enzyme capable of enzymatically converting said first amino acid sequence.

In one embodiment, the present invention contemplates a method for detecting protein-protein interactions, comprising: (a) providing a host cell comprising a detectable gene, wherein the detectable gene expresses a detectable protein when the detectable gene is activated by an amino acid sequence comprising a transcriptional activation domain; (b) providing a first chimeric gene that is capable of being expressed in said host cell, the first chimeric gene comprising a DNA sequence that encodes a first hybrid protein, the first hybrid protein comprising: (i) a DNA-binding moiety that recognizes a binding site on the detectable gene in the host cell; (ii) a reactive moiety capable of being modified through catalysis; and (iii) a catalytic moiety that is capable of catalyzing said reactive moiety; (c) providing a second chimeric gene that is capable of being expressed in the host cell, the second chimeric gene comprising a DNA sequence that encodes a second hybrid protein, the second hybrid protein comprising a transcriptional activation domain; and (d) introducing the first chimeric gene and the second chimeric gene into the host cell under conditions wherein the first hybrid protein and the second hybrid protein are expressed.

In another embodiment, the present invention contemplates that the above method comprises determining whether the detectable gene has been expressed.

In one embodiment, the present invention contemplates a compound, comprising a) a first amino acid sequence comprising at least a portion of a histone amino terminal tail, said first amino acid sequence linked to b) a second amino acid sequence comprising at least a portion of a protein kinase. In another embodiment, the

present invention contemplates that the second amino acid sequence comprises the active domain of IPL1 kinase. As with the compound embodied above, this embodiment of the present invention contemplates that, in certain embodiments, the second amino acid sequence may be catalytically inactive, that the first amino acid sequence comprises the histone H3 tail or the histone H4 tail, that the compound comprise a fusion protein, that the compound exhibits autophosphorylation, that the compound further comprises a DNA binding moiety, that the DNA binding moiety is bound to the first amino acid sequence, that the DNA binding moiety may comprise Gal4, that the second amino acid sequence comprises a detectable moiety and that said detectable moiety comprises an epitope. One embodiment of the present invention also contemplates the nucleic acid encoding the fusion protein above, an expression vector comprising that nucleic acid, a yeast transformed with that expression vector and the whole cell extract form that yeast.

In one embodiment, the present invention contemplates a compound, comprising a) a first amino acid sequence comprising at least a portion of a carboxy terminal domain, said first amino acid sequence linked to b) a second amino acid sequence comprising at least a portion of protein kinase. In another embodiment, the present invention contemplates that the second amino acid sequence comprises the active domain of KIN28 kinase. In another embodiment, the present invention contemplates that the second amino acid sequence comprises the active domain of KIN28 kinase. As with the compound embodied above this embodiment of the present invention contemplates that, if desired, the second amino acid sequence may be catalytically inactive, that the first amino acid sequence comprises a trimer of the carboxyl terminal domain, that the compound comprise a fusion protein, that the compound exhibits autophosphorylation, that the compound further comprises a DNA binding moiety, that the DNA binding moiety is bound to the first amino acid sequence, that the DNA binding moiety may comprise Gal4, that the second amino acid sequence comprises a detectable moiety and that said detectable moiety comprises an epitope. One embodiment of the present invention also contemplates the nucleic acid encoding the fusion protein above, an expression vector comprising that nucleic

13

acid, a yeast transformed with that expression vector and the whole cell extract form that yeast.

One embodiment of the present invention also contemplates a compound comprising acetylated PIASxα bound to p53.

5 **DEFINITIONS**

In order to better understand the invention, the following definitions are provided.

The terms "protein," "peptide" and "polypeptide" refer to compounds comprising amino acids joined via peptide bonds and these terms are used

10 interchangeably. A "protein," "peptide" or "polypeptide" encoded by a gene is not limited to the amino acid sequence encoded by the gene, but includes post-translational modifications of the protein. A "protein," "peptide" or "polypeptide" will also refer to a region or fragment of the named peptide.

Where the term "amino acid sequence" is recited herein to refer to an amino

15 acid sequence of a protein molecule, "amino acid sequence" and like terms, such as "polypeptide," "peptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule. Furthermore, an "amino acid sequence" can be deduced from the nucleic acid sequence encoding the protein.

20 The term "portion" when used in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino sequence minus one amino acid. The term "portion" when used in reference to a nucleotide sequence (as in "a portion of a given nucleotide sequence") refers to fragments of that nucleotide sequence. The

25 fragments may range in size from ten nucleotide residues to the entire nucleotide sequence minus one nucleotide.

"At least a portion of a histone amino terminal tail" shall be defined as a fragment of a histone amino acid tail of at least four amino acids.

"At least a portion of a histone acetyl transferase" shall be defined as a

14

fragment of a histone acetyl transferase of at least four amino acids.

"Histone H3 tail" shall be defined as the N-terminal portion of the H3 peptide. The tail is approximately 10-20 amino acids in length. The histone tail is believed to be important in transcriptional regulation.

"Histone H4 tail" shall be defined as the N-terminal portion of the H4 peptide. The tail is approximately 10-20 amino acids in length. The histone tail is believed to be important in transcriptional regulation.

"Active domain" shall be defined as the portion of a molecule that has functional properties such as, but not limited to, catalytic and enzymatic properties.

The term "chimera" when used in reference to a polypeptide refers to the expression product of two or more coding sequences obtained from different genes, that have been cloned together and that, after translation, act as a single polypeptide sequence. Chimeric polypeptides are also referred to as "hybrid" polypeptides. The coding sequences includes those obtained from the same or from different species of organisms. Chimeric peptides are produced form "chimeric genes."

The term "fusion" when used in reference to a polypeptide refers to a chimeric protein containing a protein of interest joined to an exogenous protein fragment (the fusion partner). The fusion partner may serve various functions, including enhancement of solubility of the polypeptide of interest, as well as providing an "affinity tag" to allow purification of the recombinant fusion polypeptide from a host cell or from a supernatant or from both. If desired, the fusion partner may be removed from the protein of interest after or during purification.

The term "homolog" or "homologous" when used in reference to a polypeptide refers to a high degree of sequence identity between two polypeptides, or to a high degree of similarity between the three-dimensional structure or to a high degree of similarity between the active site and the mechanism of action. In a preferred embodiment, a homolog has a greater than 60 % sequence identity, and more preferably greater than 75 % sequence identity, and still more preferably greater than 90 % sequence identity, with a reference sequence.

As applied to polypeptides, the term "substantial identity" means that two

15

peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity or more (*e.g.*, 99 percent sequence identity). Preferably, residue positions

5   which are not identical differ by conservative amino acid substitutions.

The terms "variant" and "mutant" when used in reference to a polypeptide refer to an amino acid sequence that differs by one or more amino acids from another, usually related polypeptide. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties. One type of

10   conservative amino acid substitutions refers to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids

15   having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

20   More rarely, a variant may have "non-conservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Similar minor variations may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological activity may be found using computer programs well known in

25   the art, for example, DNAStar software. Variants can be tested in functional assays. Preferred variants have less than 10 %, and preferably less than 5 %, and still more preferably less than 2 % changes (whether substitutions, deletions, and so on).

The term "domain" when used in reference to a polypeptide refers to a subsection of the polypeptide which possesses a unique structural and/or functional

30   characteristic; typically, this characteristic is similar across diverse polypeptides. The

subsection typically comprises contiguous amino acids, although it may also comprise amino acids which act in concert or which are in close proximity due to folding or other configurations.

The term "gene" refers to a nucleic acid (*e.g.*, DNA sequence, RNA sequence or nucleotide sequence) sequence that comprises coding sequences necessary for the production of an RNA, or a polypeptide or its precursor (*e.g.*, proinsulin). A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, etc.) of the polypeptide are retained. The term "portion" when used in reference to a gene refers to fragments of that gene. The fragments may range in size from a few nucleotides to the entire gene sequence minus one nucleotide. Thus, "a nucleotide comprising at least a portion of a gene" may comprise fragments of the gene or the entire gene.

The term "gene" also encompasses the coding regions of a structural gene and includes sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide. A "translation product" of a DNA sequence is the peptide sequence generated via from the mRNA encoded by the DNA.

In addition to containing introns, genomic forms of a gene may also include

17

sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

The term "heterologous" when used in reference to a gene refers to a gene encoding a peptide that is not in its natural environment (*i.e.*, has been altered by the hand of man). For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.g.*, mutated, added in multiple copies, linked to a non-native promoter or enhancer sequence, etc.). Heterologous genes may comprise gene sequences that comprise cDNA forms of a gene; the cDNA sequences may be expressed in either a sense (to produce mRNA) or anti-sense orientation (to produce an anti-sense RNA transcript that is complementary to the mRNA transcript). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to nucleotide sequences comprising regulatory elements such as promoters that are not found naturally associated with the gene for the protein encoded by the heterologous gene or with gene sequences in the chromosome, or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

The term "nucleotide sequence of interest" or "nucleic acid sequence of interest" refers to any nucleotide sequence (*e.g.*, RNA or DNA), the manipulation of which may be deemed desirable for any reason (*e.g.*, treat disease, confer improved qualities, *etc.*), by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (*e.g.*, reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, *etc.*), and non-coding regulatory sequences which do not encode an mRNA or protein product (*e.g.*, promoter sequence, polyadenylation sequence, termination sequence, enhancer

18

sequence, *etc.*).

The term "structural" when used in reference to a gene or to a nucleotide or nucleic acid sequence refers to a gene or a nucleotide or nucleic acid sequence whose ultimate expression product is a protein (such as an enzyme or a structural protein), an

5   rRNA, an sRNA, a tRNA, etc.

The terms "oligonucleotide" or "polynucleotide" or "nucleotide sequence" or "nucleic acid sequence" refer to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three, and usually more than ten. The exact size will depend on many factors, which in turn depends on the

10   ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof.

The terms "an oligonucleotide having a nucleotide sequence encoding a gene" or "a nucleic acid sequence encoding" a specified polypeptide refer to a nucleic acid

15   sequence comprising the coding region of a gene or in other words the nucleic acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions,

20   polyadenylation signals, *etc.* may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals,

25   *etc.* or a combination of both endogenous and exogenous control elements.

The term "recombinant" when made in reference to a nucleic acid molecule refers to a nucleic acid molecule which is comprised of segments of nucleic acid joined together by means of molecular biological techniques. The term "recombinant" when made in reference to a protein or a polypeptide refers to a protein molecule

30   which is expressed using a recombinant nucleic acid molecule.

The terms "complementary" and "complementarity" refer to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "homology" when used in relation to nucleic acids refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). "Sequence identity" refers to a measure of relatedness between two or more nucleic acids or proteins, and is given as a percentage with reference to the total comparison length. The identity calculation takes into account those nucleotide or amino acid residues that are identical and in the same relative positions in their respective larger sequences. Calculations of identity may be performed by algorithms contained within computer programs such as "GAP" (Genetics Computer Group, Madison, Wis.) and "ALIGN" (DNAStar, Madison, Wis.). A partially complementary sequence is one that at least partially inhibits (or competes with) a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a sequence which is completely homologous to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of

20

complementarity (*e.g.*, less than about 30 % identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence," "sequence identity," "percentage of sequence identity" and "substantial identity." A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA sequence given in a sequence listing or may comprise a complete gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (*i.e.*, a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window," as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman [Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981)] by the homology alignment algorithm of Needleman and Wunsch [Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970)], by the search for similarity method of Pearson and Lipman [Pearson and Lipman, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:2444 (1988)], by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and

21

the best alignment (*i.e.*, resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected. The term "sequence identity" means that two polynucleotide sequences are identical (*i.e.*, on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence, for example, as a segment of the full-length sequences of the compositions claimed in the present invention.

The term "substantially homologous" when used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low to high stringency as described above.

The term "substantially homologous" when used in reference to a single-stranded nucleic acid sequence refers to any probe that can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low to high stringency as described above.

The term "hybridization" refers to the pairing of complementary nucleic acids.

22

Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains

5   pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

The term "T_m" refers to the "melting temperature" of a nucleic acid. The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for

10   calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% \ G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (*See e.g.*, Anderson and Young, Quantitative Filter Hybridization, in *Nucleic Acid Hybridization* [1985]). Other references include more sophisticated computations that

15   take structural as well as sequence characteristics into account for the calculation of T_m.

The term "stringency" refers to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid

20   base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "low" stringency are often required with nucleic acids that are derived from organisms that are genetically diverse, as the frequency of complementary sequences is usually less.

"Low stringency conditions" when used in reference to nucleic acid

25   hybridization comprise conditions equivalent to binding or hybridization at 42 °C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4(H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1 % SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 μg/ml denatured salmon sperm DNA followed by washing in a

30   solution comprising 5X SSPE, 0.1 % SDS at 42 °C when a probe of about 500

23

nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42 $^{\circ}$C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH$_2$PO$_4$(H$_2$O and 1.85 g/l

5 EDTA, pH adjusted to 7.4 with NaOH), 0.5 % SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0 % SDS at 42 $^{\circ}$C when a probe of about 500 nucleotides in length is employed.

"High stringency conditions" when used in reference to nucleic acid

10 hybridization comprise conditions equivalent to binding or hybridization at 42 $^{\circ}$C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH$_2$PO$_4$(H$_2$O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5 % SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0 % SDS at 42 $^{\circ}$C when a probe of about 500 nucleotides in length is

15 employed.

It is well known that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts

20 and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (e.g., increasing the temperature of

25 the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.).

The term "wild-type" when made in reference to a gene refers to a gene that has the characteristics of a gene isolated from a naturally occurring source. The term "wild-type" when made in reference to a gene product refers to a gene product that has

30 the characteristics of a gene product isolated from a naturally occurring source. The

term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is

5   naturally-occurring. A wild-type gene is frequently that gene which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" when made in reference to a gene or to a gene product refers, respectively, to a gene or to a gene product which displays modifications in sequence and/or functional properties (*i.e.*,

10   altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

Thus, the terms "variant" and "mutant" when used in reference to a nucleotide

15   sequence refer to an nucleic acid sequence that differs by one or more nucleotides from another, usually related nucleotide acid sequence. A "variation" is a difference between two different nucleotide sequences; typically, one sequence is a reference sequence.

The term "polymorphic locus" refers to a genetic locus present in a population

20   that shows variation between members of the population (*i.e.*, the most common allele has a frequency of less than 0.95). Thus, "polymorphism" refers to the existence of a character in two or more variant forms in a population. A "single nucleotide polymorphism" (or SNP) refers a genetic locus of a single base which may be occupied by one of at least two different nucleotides. In contrast, a "monomorphic

25   locus" refers to a genetic locus at which little or no variations are seen between members of the population (generally taken to be a locus at which the most common allele exceeds a frequency of 0.95 in the gene pool of the population).

A "frameshift mutation" refers to a mutation in a nucleotide sequence, usually resulting from insertion or deletion of a single nucleotide (or two or four nucleotides)

30   which results in a change in the correct reading frame of a structural DNA sequence

25

encoding a protein. The altered reading frame usually results in the translated amino-acid sequence being changed or truncated.

A "splice mutation" refers to any mutation that affects gene expression by affecting correct RNA splicing. Splicing mutation may be due to mutations at intron-exon boundaries which alter splice sites.

The term "detection assay" refers to an assay for detecting the presence or absence of a sequence or a variant nucleic acid sequence (*e.g.*, mutation or polymorphism in a given allele of a particular gene or for detecting the presence or absence of a particular protein or the structure or activity or effect of a particular protein or for detecting the presence or absence of a variant of a particular protein.

The term "antisense" refers to a deoxyribonucleotide sequence whose sequence of deoxyribonucleotide residues is in reverse 5' to 3' orientation in relation to the sequence of deoxyribonucleotide residues in a sense strand of a DNA duplex. A "sense strand" of a DNA duplex refers to a strand in a DNA duplex which is transcribed by a cell in its natural state into a "sense mRNA." Thus an "antisense" sequence is a sequence having the same sequence as the non-coding strand in a DNA duplex. The term "antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene by interfering with the processing, transport and/or translation of its primary transcript or mRNA. The complementarity of an antisense RNA may be with any part of the specific gene transcript, *i.e.*, at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. In addition, as used herein, antisense RNA may contain regions of ribozyme sequences that increase the efficacy of antisense RNA to block gene expression. "Ribozyme" refers to a catalytic RNA and includes sequence-specific endoribonucleases. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of preventing the expression of the target protein.

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template).

26

Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Qb replicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.*, Proc. Natl. Acad. Sci. USA, 69:3038 [1972]). Other nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlain *et al.*, Nature, 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace, Genomics, 4:560 [1989]). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), *PCR Technology*, Stockton Press [1989]).

The term "amplifiable nucleic acid" refers to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

The term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target" (defined below). In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid

contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

The term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

The term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

The term "target," when used in reference to the polymerase chain reaction, refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target

28

sequence.

The term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188, that describe a method for increasing the concentration of a segment of a target sequence in a mixture of

5   genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To

10   effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle";

15   there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase

20   chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified."

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*,

25   hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created

30   by the PCR process itself are, themselves, efficient templates for subsequent PCR

29

amplifications.

The terms "PCR product," "PCR fragment," and "amplification product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences.

The term "amplification reagents" refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction vessel (test tube, microwell, etc.).

The term "reverse-transcriptase" or "RT-PCR" refers to a type of PCR where the starting material is mRNA. The starting mRNA is enzymatically converted to complementary DNA or "cDNA" using a reverse transcriptase enzyme. The cDNA is then used as a "template" for a "PCR" reaction

The term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (e.g., mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (i.e., via the enzymatic action of an RNA polymerase), and into protein, through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products (i.e., RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. Molecules (e.g., transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and "repressors," respectively.

The terms "in operable combination", "in operable order" and "operably linked" refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

The term "regulatory element" refers to a genetic element which controls some

30

aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, *etc.*

5   Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis, *et al.*, Science 236:1237, 1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect,

10   mammalian and plant cells. Promoter and enhancer elements have also been isolated from viruses and analogous control elements, such as promoters, are also found in prokaryotes. The selection of a particular promoter and enhancer depends on the cell type used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types

15   (*for review, see* Voss, *et al.*, Trends Biochem. Sci., 11:287, 1986; and Maniatis, *et al.*, *supra* 1987).

The terms "promoter element," "promoter," or "promoter sequence" refer to a DNA sequence that is located at the 5' end (*i.e.* precedes) of the coding region of a DNA polymer. The location of most promoters known in nature precedes the

20   transcribed region. The promoter functions as a switch, activating the expression of a gene. If the gene is activated, it is said to be transcribed, or participating in transcription. Transcription involves the synthesis of mRNA from the gene. The promoter, therefore, serves as a transcriptional regulatory element and also provides a site for initiation of transcription of the gene into mRNA.

25   The term "regulatory region" refers to a gene's 5' transcribed but untranslated regions, located immediately downstream from the promoter and ending just prior to the translational start of the gene.

The term "promoter region" refers to the region immediately upstream of the coding region of a DNA polymer, and is typically between about 500 bp and 4 kb in

30   length, and is preferably about 1 to 1.5 kb in length.

31

Promoters may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue in the relative absence of expression of the same nucleotide sequence of interest in a different

5   type of tissue. Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the

10   activity of a protein encoded by the reporter gene) in different tissues of the transgenic animal. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected. The term "cell type specific" as applied to a promoter refers to a

15   promoter which is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a

20   single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, e.g., immunohistochemical staining. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleotide sequence of interest whose expression is controlled by the promoter. A labeled (e.g., peroxidase

25   conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy.

Promoters may be constitutive or inducible. The term "constitutive" when made in reference to a promoter means that the promoter is capable of directing

30   transcription of an operably linked nucleic acid sequence in the absence of a stimulus

(*e.g.*, heat shock, chemicals, light, etc.). Typically, constitutive promoters are capable of directing expression of a transgene in substantially any cell and any tissue.

In contrast, an "inducible" promoter is one which is capable of directing a level of transcription of an operably linked nucleic acid sequence in the presence of a

5   stimulus (*e.g.*, heat shock, chemicals, light, *etc.*) which is different from the level of transcription of the operably linked nucleic acid sequence in the absence of the stimulus.

The term "regulatory element" refers to a genetic element that controls some aspect of the expression of nucleic acid sequence(s). For example, a promoter is a

10   regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etc.

The enhancer and/or promoter may be "endogenous" or "exogenous" or "heterologous." An "endogenous" enhancer or promoter is one that is naturally linked

15   with a given gene in the genome. An "exogenous" or "heterologous" enhancer or promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (*i.e.*, molecular biological techniques) such that transcription of the gene is directed by the linked enhancer or promoter. For example, an endogenous promoter in operable combination with a first gene can be isolated, removed, and placed in

20   operable combination with a second gene, thereby making it a "heterologous promoter" in operable combination with the second gene. A variety of such combinations are contemplated (*e.g.*, the first and second genes can be from the same species, or from different species).

The term "naturally linked" or "naturally located" when used in reference to the

25   relative positions of nucleic acid sequences means that the nucleic acid sequences exist in nature in the relative positions.

The presence of "splicing signals" on an expression vector often results in higher levels of expression of the recombinant transcript in eukaryotic host cells. Splicing signals mediate the removal of introns from the primary RNA transcript and

30   consist of a splice donor and acceptor site (Sambrook, *et al.*, *Molecular Cloning: A*

33

Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York [1989] pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term "poly(A) site" or "poly(A) sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable, as transcripts lacking a poly(A) tail are unstable and are rapidly degraded. The poly(A) signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous poly(A) signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly(A) signal is one which has been isolated from one gene and positioned 3' to another gene. A commonly used heterologous poly(A) signal is the SV40 poly(A) signal. The SV40 poly(A) signal is contained on a 237 bp *Bam*HI/*Bcl*I restriction fragment and directs both termination and polyadenylation (Sambrook, *supra*, at 16.6-16.7).

The term "vector" refers to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vehicle" is sometimes used interchangeably with "vector."

The terms "expression vector" or "expression cassette" refer to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The term "transfection" refers to the introduction of foreign DNA into cells. Transfection may be accomplished by a variety of means known to the art including

34

calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, glass beads, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, viral infection, biolistics (*i.e.*, particle bombardment) and the like.

5   The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell that has stably integrated foreign DNA into the genomic DNA.

The term "transient transfection" or "transiently transfected" refers to the

10   introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells that have taken up

15   foreign DNA but have failed to integrate this DNA.

The term "stable expression" means the expression of an exogenous sequence wherein the transfected sequences has been integrated into the genome.

The term "transient expression" means the expression of an exogenous sequence wherein the transfected sequences has failed to integrate into the genome.

20   The term "calcium phosphate co-precipitation" refers to a technique for the introduction of nucleic acids into a cell. The uptake of nucleic acids by cells is enhanced when the nucleic acid is presented as a calcium phosphate-nucleic acid co-precipitate. The original technique of Graham and van der Eb (Graham and van der Eb, Virol., 52:456 [1973]), has been modified by several groups to optimize

25   conditions for particular types of cells. The art is well aware of these numerous modifications.

The terms "infecting" and "infection" when used with a bacterium refer to co-incubation of a target biological sample, (*e.g.*, cell, tissue, *etc.*) with the bacterium under conditions such that nucleic acid sequences contained within the bacterium are

30   introduced into one or more cells of the target biological sample.

35

The terms "bombarding, "bombardment," and "biolistic bombardment" refer to the process of accelerating particles towards a target biological sample (*e.g.*, cell, tissue, *etc.*) to effect wounding of the cell membrane of a cell in the target biological sample and/or entry of the particles into the target biological sample. Methods for

5   biolistic bombardment are known in the art (*e.g.*, U.S. Patent No. 5,584,807, the contents of which are incorporated herein by reference), and are commercially available (*e.g.*, the helium gas-driven microprojectile accelerator (PDS-1000/He, BioRad).

The term "transgene" refers to a foreign gene that is placed into an organism by

10   the process of transfection. The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) that is introduced into the genome of an organism by experimental manipulations and may include gene sequences found in that organism so long as the introduced gene does not reside in the same location as does the naturally-occurring gene.

15   The term "transgenic" when used in reference to a host cell or an organism refers to a host cell or an organism that contains at least one heterologous or foreign gene in the host cell or in one or more of cells of the organism.

The term "host cell" refers to any cell capable of replicating and/or transcribing and/or translating a heterologous gene. Thus, a "host cell" refers to any eukaryotic or

20   prokaryotic cell (*e.g.*, bacterial cells such as *E. coli*, yeast cells, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells), whether located *in vitro* or *in vivo*. For example, host cells may be located in a transgenic animal.

The terms "transformants" or "transformed cells" include the primary transformed cell and cultures derived from that cell without regard to the number of

25   transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

The term "selectable marker" refers to a gene which encodes an enzyme having

30   an activity that confers resistance to an antibiotic or drug upon the cell in which the

36

selectable marker is expressed, or which confers expression of a trait which can be detected (*e.g.*, luminescence or fluorescence). Selectable markers may be "positive" or "negative." Examples of positive selectable markers include the neomycin phosphotrasferase (NPTII) gene which confers resistance to G418 and to kanamycin,

5   and the bacterial hygromycin phosphotransferase gene (*hyg*), which confers resistance to the antibiotic hygromycin. Negative selectable markers encode an enzymatic activity whose expression is cytotoxic to the cell when grown in an appropriate selective medium. For example, the HSV-*tk* gene is commonly used as a negative selectable marker. Expression of the HSV-*tk* gene in cells grown in the presence of

10   gancyclovir or acyclovir is cytotoxic; thus, growth of cells in selective medium containing gancyclovir or acyclovir selects against cells capable of expressing a functional HSV TK enzyme.

The term "reporter gene" refers to a gene encoding a protein that may be assayed. Examples of reporter genes include, but are not limited to, luciferase (*See,*

15   *e.g.*, deWet *et al.*, Mol. Cell. Biol. 7:725 [1987] and U.S. Pat Nos.,6,074,859; 5,976,796; 5,674,713; and 5,618,682; all of which are incorporated herein by reference), green fluorescent protein (*e.g.*, GenBank Accession Number U43284; a number of GFP variants are commercially available from CLONTECH Laboratories, Palo Alto, CA), chloramphenicol acetyltransferase, ß-galactosidase, alkaline

20   phosphatase, and horse radish peroxidase.

The term "overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. The term "cosuppression" refers to the expression of a foreign gene which has substantial homology to an endogenous gene resulting in the suppression of

25   expression of both the foreign and the endogenous gene. As used herein, the term "altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

The terms "Southern blot analysis" and "Southern blot" and "Southern" refer to the analysis of DNA on agarose or acrylamide gels in which DNA is separated or

30   fragmented according to size followed by transfer of the DNA from the gel to a solid

37

support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then exposed to a labeled probe to detect DNA species complementary to the probe used. The DNA may be cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during

5   transfer to the solid support. Southern blots are a standard tool of molecular biologists (J. Sambrook *et al.* [1989] *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9.31-9.58).

The term "Northern blot analysis" and "Northern blot" and "Northern" refer to the analysis of RNA by electrophoresis of RNA on agarose gels to fractionate the

10   RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists (J. Sambrook, *et al.* [1989] *supra*, pp 7.39-7.52).

15   The terms "Western blot analysis" and "Western blot" and "Western" refers to the analysis of protein(s) (or polypeptides) immobilized onto a support such as nitrocellulose or a membrane. A mixture comprising at least one protein is first separated on an acrylamide gel, and the separated proteins are then transferred from the gel to a solid support, such as nitrocellulose or a nylon membrane. The

20   immobilized proteins are exposed to at least one antibody with reactivity against at least one antigen of interest. The bound antibodies may be detected by various methods, including the use of radiolabeled antibodies.

The term "antigenic determinant" refers to that portion of an antigen that makes contact with a particular antibody (*i.e.,* an epitope). When a protein or fragment of a

25   protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies that bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (*i.e.,* the "immunogen" used to elicit the immune response) for binding to an antibody.

30   The term "isolated" when used in relation to a nucleic acid, as in "an isolated

38

oligonucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids, such as DNA and RNA, are found in the state they exist in nature. Examples of non-isolated nucleic acids include: a given DNA sequence (e.g., a gene) found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, isolated nucleic acid encoding a particular protein includes, by way of example, such nucleic acid in cells ordinarily expressing the protein, where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded).

The term "purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated. An "isolated nucleic acid sequence" may therefore be a purified nucleic acid sequence. "Substantially purified" molecules are at least 60 % free, preferably at least 75 % free, and more preferably at least 90 % free from other components with which they are naturally associated. As used herein, the term "purified" or "to purify" also refer to the removal of contaminants from a sample. The removal of contaminating proteins results in an increase in the percent of polypeptide of interest in the sample. In another example, recombinant polypeptides are expressed in plant, bacterial, yeast, or mammalian host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

The term "composition comprising" a given polynucleotide sequence or

polypeptide refers broadly to any composition containing the given polynucleotide sequence or polypeptide. The composition may comprise an aqueous solution. In one embodiment, polynucleotide sequences are typically employed in an aqueous solution containing salts (*e.g.*, NaCl), detergents (*e.g.*, SDS), and other components (*e.g.*, Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The term "test compound" refers to any chemical entity, pharmaceutical, drug, and the like that can be used to treat or prevent a disease, illness, sickness, or disorder of bodily function, or otherwise alter the physiological or cellular status of a sample. Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods of the present invention. A "known therapeutic compound" refers to a therapeutic compound that has been shown (*e.g.*, through animal trials) to be effective in such treatment or prevention.

As used herein, the term "response," when used in reference to an assay, refers to the generation of a detectable signal (*e.g.*, accumulation of reporter protein, increase in ion concentration, accumulation of a detectable chemical product).

The terms "sample" and "source" are used in their broadest sense. In one sense they can refer to a animal cell or tissue. In another sense, they is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from plants or animals (including humans) and encompass fluids, solids, tissues, and gases. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

The term "immunohistochemical assay" is defined as an assay that comprises peptides (*e.g.*, antibodies) that recognized antigenic determinants (*e.g.*, epitopes). The peptides are linked either directly or indirectly to other peptides or other compounds (*e.g.*, fluorescent peptides or chemicals, enzymes and the like) that give a detectable signal in the given assay system. An example of an immunohistochemical assay would be an ELISA assay.

40

"Portion" of a peptide shall be defined as a sequence of at least 10 amino acids up to the total length of the peptide less one amino acid. In a preferred embodiment, portion shall include Lys9 and/or Lys14 amino acid of the histone H3 tail.

"Linked," in regards to the peptides of the present invention, shall be defined as peptides or peptide portions that are connected via peptide bonds or via chemical bonds.

"Autoacetylation" shall be defined an enzymatic compound (*e.g.*, a protein or peptide) that has the ability to acetylate an amino acid residue on same compound. In other words, autoacetylation is a form of autocatalysis.

"DNA binding moiety" shall be defined as a portion of a molecule that has the ability to bind DNA. For example, the Gal4 DNA binding domain is a DNA binding moiety.

"Detectable moiety" shall be defined as a portion of a molecule that can be readably detected by standard biochemical means. The HA moiety used in the present invention is a example of a detectable moiety.

"Epitope" shall be defined as a site on a molecule against which an antibody will be produced and to which it will bind.

"Fragment"shall be defined as a portion of, for example, a peptide, protein or nucleic acid.

"Enzyme" shall be defined as a catalyst (*e.g.*, a peptide or protein or fragment thereof) that catalyses reactions (*e.g.*, chemical or biochemical reactions) of other substances (*e.g.*, proteins, etc.) without itself being destroyed or altered upon completion of the reactions.

"Enzymatically converted" shall be defined as a substance that has been acted upon by a catalyst.

"Transcriptional activation domain" shall be defined as a nucleotide sequence that, when activated (for example, by a transcription factor) initiated the transcription of a sequence of DNA. For example, see U.S. Patent Nos. 6,271,341; 6,133,027; 5,750,667 and 6,114,111 which are incorporated herein by reference.

"In sufficient proximity" shall be defined as, for example, being close enough

41

to exert an effect on something else. For example, a transcription factor is "in sufficient proximity" to a transcriptional activation domain when it can initiate transcription. Additionally, "in sufficient proximity" as it relates to Yeast two-hybrid systems, is defined in U.S. Patent Nos. 5,667,973; 5,468,614 and 5,283,173 which are incorporated herein by reference.

"Expressed to a degree greater than" shall be defined as expression (of, for example, a gene) to a level higher (e.g., by at least approximately 10 percent higher) than another gene or the same gene in another system. In other words, it shall mean that a difference in expression is detectable by at least approximately 10 percent over background and preferably over 20 percent of more.

DESCRIPTION OF FIGURES

Figure 1 shows the design of the AC/2H system. A protein (A) is fused to either an active protein-modifying enzyme (B) or a catalytically inactivated mutant form of B (B') to create an autocatalytic hybrid protein (the wildtype B fusion) or an otherwise identical but unmodified hybrid protein (i.e. the mutant B' fusion). The DNA fragments encoding the A-B and A-B' hybrids are ligated in-frame to module C for a two-hybrid bait, and module D for easy purification and characterization of the final autocatalytic hybrid proteins. Two more controls composed of C-B-D and C-B'-D are created similarly. The curved arrow indicates the autocatalysis of the enzyme-substrate fusion protein (denoted by the "flag" on the protein A).

Figure 2 shows the application of the concept of autocatalysis to the Three-Hybrid system to identify protein-RNA interactions that require a specific modification of the RNA molecule. The module A is a known RNA binding protein that binds module D. Module A is fused in-framed to the DNA binding module C and an RNA modifying enzyme B. The module D is the known interaction partner of protein A. The module E is the RNA molecule of interest that is modified by the enzyme B. When the C-A-B hybrid protein and the D-E hybrid RNA are expressed in the same cell, the C-A-B hybrid protein recruits the D-E hybrid RNA to the promoter region and the enzyme B is thus able to modify the target RNA module E (indicated

by the "flag"). The presence of the modified E at the promoter will recruit its interaction proteins, when fused to the activation domain, and activates the transcription from the nearby reporter gene. In the control strain, the enzyme B is replaced with the catalytically inactive enzyme B' so that the modification of the RNA module E is no longer possible. The C-A-B' and D-E hybrids will be the negative control as used in Figure 1. Hence, protein-RNA interactions that require the target RNA to be modified can be detected.

Figure 3 shows the use of the autocatalysis design in identifying protein-DNA interactions that require a specific modification of the DNA element. The modules C is a sequence-specific DNA binding protein that binds the element D. The module B is a DNA modifying enzyme whereas the element E is a DNA sequence containing the modification target for the enzyme B. The D-E fusion is inserted in front of the reporter gene (not shown). When the C-B hybrid protein is expressed in the strain bearing D-E-reporter gene, the enzyme B is brought to the promoter via C-D interaction and modifies the E element. If a protein that is able to interact with the modified E element but not the unmodified E, expression from the reporter gene will be detected. On the other hand, if the enzyme B is substituted with the mutant B', no modification of E will be yielded; interactions involving the E element but do not require its modification by B can be sorted out by the B' control hybrid.

Figure 4 shows the design of the autoacetylated H3/H4-Gcn5 baits for the Yeast Two-Hybrid screening. The histone tails H3 and H4 are individually fused to the catalytic domain of a histone acetyltransferase Gcn5, the Gal4 DNA binding domain (GDBD), and the HA epitope tag to create plasmid constructs pDG1-4 (low-copy yeast vectors) and pDG5-8 (high-copy yeast vectors). When wildtype Gcn5 is included in the fusion, autocatalysis results in the acetylation of the fused histone H3 or H4 (the "lollipop") (see, for example, Figure 12, pDG1 [SEQ ID NOS: 1 and 15]; Figure 13, pDG2 [SEQ ID NOS: 2 and 16]; Figure 16, pDG5 [SEQ ID NOS: 5 and 19] and; Figure 17, pDG6 [SEQ ID NOS: 6 and 20]). The mutant Gcn5 F221A, marked by the "thunderbolt", fails to catalyze the histone acetylation (see, for example, Figure 14, pDG3 [SEQ ID NOS: 3 and 17]; Figure 15, pDG4 [SEQ ID NOS: 4 and

43

18]; Figure 18, pDG7 [SEQ ID NOS: 7 and 21] and; Figure 19, pDG8 [SEQ ID NOS: 8 and 22]). In the yeast strain where the GDBD-H3 or H4-Gcn5-HA is co-expressed with a corresponding acetylated histone binding protein that is fused to a transcriptional activation domain, the reporter gene under the control of the enhancer element UASgal will be activated due to the interaction between the autoacetylated histone bait and the specific prey protein. On the other hand, such interactions will not be seen if the mutant Gcn5 F221A is part of the autocatalytic bait fusion.

Figures 5A, B and C show autoacetylation of H3-Gcn5 and H4-Gcn5 within the context of the Ras Recruitment System. To test the feasibility of the autocatalysis in two-hybrid tests, the H3-Gcn5 (wildtype or F221A mutant) was fused in-frame to the glutathione (GST) and the Ras protein. The H3-GST-Gcn5-Ras fusion was expressed in bacteria (A) and yeast (B) prior to purification for western analyses. The purified fusion proteins were then resolved by SDS-PAGE and analyzed by an antibody specific for histone H3 acetylation. In addition, histone H4 was fused to Gcn5 in parallel experiments and were purified from yeast and probed with an antibody against acetylated H4 (C). In all cases, the wildtype Gcn5 fusion leads to autoacetylation of H3 and H4, in both bacteria and yeast host strains, whereas the mutant Gcn5 fusion failed to do so, providing an ideal negative control for acetylation-dependent protein-protein interactions. Furthermore, the H3-Gcn5 protein expressed and purified from bacteria was also subjected to western analyses using an antibody specific for unacetylated H3 (A, middle panel). The very weak signal seen associated with the wildtype Gcn5 fusion indicates that the autocatalysis is very efficient, such that the residual unacetylated H3 can not be detected effectively by this antibody.

Figure 6 shows the autoacetylation of H3-Gcn5 in the classical yeast two-hybrid setting. The design of the autocatalytic baits is shown in Figure 4. The yeast proteins were immunoprecipitated by the anti-HA antibody, followed by SDS-PAGE and western analyses using antibodies against HA (as a loading control, left panel), and against the acetylated H3 (to see the acetylation status, right panel). The positions of the fusion proteins of interest are marked by the arrows.

Figure 7 shows a previously reported acetylated histone interaction can be

44

detected by the AC/Y2H method. The bromodomain of PCAF was fused to the Gal4 activation domain (AD). The PCAF-AD and the AD-only vectors were transformed into the bait-containing yeast strains. The UASgal-lacZ reporter gene expression was assessed by measuring the b-galactosidase activity from log-phase cells (Y axis:

5   units/mg protein/min). Stronger lacZ expression indicates protein-protein interactions.

Figure 8 show the identification of acetylated histone H3- and H4-binding proteins by the AC/2H system: high-throughput method. The composition of the baits are shown on the top (3, H3; 4, H4; w, wildtype Gcn5; m, F221A Gcn5); the AD fusion preys are listed on the right. Three media were used: -His medium is more

10   sensitive and allows weaker interaction to be detected; -Ade plate reveals stronger interactions. The Rpd3 + H3-Gcn5 (wt) is blown up to show the weak/transient interaction revealed by the -His medium. Note that the H4-Gcn5 bait activates moderate transcription so that all strains containing this bait are able to grow on -His plate. However, on -Ade plate, the H4-Gcn5 bait does not induce high enough ADE2

15   reporter expression. Therefore, strong Y2H interactions are detectable in this medium.

Figure 9 shows the identification of acetylated histone H3- and H4-binding proteins by the AC/2H system: AD library screening. Two putative AHBPs (acetylated histone binding proteins) were identified by AD library screening using the GDBD-H3-Gcn5-HA bait. Different bait constructs are listed on the right (GDBD and

20   HA are omitted from the legends). The growth on the -Ade plate (left) was assessed. Clone 5 and, to a lesser degree, clone 1 showed obvious growth on the -Ade plate only when the H3-Gcn5 (wt) bait was present, indicating the exclusive interaction with an acetylated H3. On the other hand, clone 6 caused robust growth whenever the wildtype Gcn5 was included in the bait, suggesting 1) that the wildtype Gcn5 interacts

25   with this protein, or 2) that the autoacetylated H3 bait interacts with this prey, or 3) that the enzymatically active Gcn5, while tethered to the promoter region, acetylates nearby histones and/or other protein factors that act as the interacting partner for the prey. DNA sequencing analyses showed that these three candidates are a novel peptide (clone 5), Rpm2 (clone 1), and Cin8 (clone 6).

30   Figures 10A and B show the autoacetylation of p53 by Gcn5. (A) Schematic

45

drawing of the p53 autoacetylation constructs. (B) p53 fusion proteins were expressed in yeast, immunoprecipitated by antibodies against HA or against acetylated Lys320 (K320.Ac), and then tested by western analyses to quantify the relative abundance and the status of K320 acetylation. It is clear that p53 is acetylated when fused to the wildtype Gcn5, but not by the mutant Gcn5 in the autocatalysis context. Two independent yeast colonies bearing these two autocatalysis baits were analyzed in parallel.

Figure 11 shows the identification of proteins that interact with acetylated or unacetylated p53 protein. The GDBD-p53-Gcn5(wt)-HA (pMK485, Figure 24, SEQ ID NOS: 13 and 27), GDBD-p53-Gcn5(mutant)-HA (pMK486, Figure 25, SEQ ID NOS: 14 and 28), as well GDBD-Gcn5(wt)-HA were used as the baits in the yeast two-hybrid method to screen for human proteins that interact with specific p53 species. The yeast transformants bearing the activation domain-human cDNA fusion constructs and are able to active the ADE2 reporter gene became ADE+ and were tested for their ability to survive in the absence of adenine (SC -Leu -Trp -Ade plates) in one of the three baits mentioned above. Three classes of the candidates were identified: Class I represents those that only interact with the wildtype Gcn5 fusion of p53 (i.e. the acetylated p53 protein); Class II represents those that interact with both wildtype and mutant Gcn5 fusion of p53 (*i.e.*, p53 interactors independent of the acetylation status); Class III represent those that interact with both wildtype Gcn5-p53 fusion as well as the wildtype Gcn5 alone.

Figure 12 shows the nucleotide sequence (SEQ ID NO: 1) and amino acid sequence (SEQ ID NO: 15) of the coding region of the plasmid pDG1. pDG1 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, histone H3 (amino acids 1-59), Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 13 shows the nucleotide sequence (SEQ ID NO: 2) and amino acid sequence (SEQ ID NO: 16) of the coding region of the plasmid pDG2. pDG2 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, histone H3 (amino acids 1-59), Gcn5 F221A mutant allele (amino acids 18-252),

46

and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 14 shows the nucleotide sequence (SEQ ID NO: 3) and amino acid sequence (SEQ ID NO: 17) of the coding region of the plasmid pDG3. pDG3 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, histone H4 (amino acids 1-29), Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 15 shows the nucleotide sequence (SEQ ID NO: 4) and amino acid sequence (SEQ ID NO: 18) of the coding region of the plasmid pDG4. pDG4 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, histone H4 (amino acids 1-29), Gcn5 F221A mutant allele (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 16 shows the nucleotide sequence (SEQ ID NO: 5) and amino acid sequence (SEQ ID NO: 19) of the coding region of the plasmid pDG5. pDG5 is a high-copy LEU2 yeast vector (YEplac181) containing Gal4 DNA binding domain, histone H3 (amino acids 1-59), Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 17 shows the nucleotide sequence (SEQ ID NO: 6) and amino acid sequence (SEQ ID NO: 20) of the coding region of the plasmid pDG6. pDG6 is a high-copy LEU2 yeast vector (YEplac181) containing Gal4 DNA binding domain, histone H3 (amino acids 1-59), Gcn5 F221A mutant allele (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 18 shows the nucleotide sequence (SEQ ID NO: 7) and amino acid sequence (SEQ ID NO: 21) of the coding region of the plasmid pDG7. pDG7 is a high-copy LEU2 yeast vector (YEplac181) containing Gal4 DNA binding domain, histone H4 (amino acids 1-29), Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 19 shows the nucleotide sequence (SEQ ID NO: 8) and amino acid sequence (SEQ ID NO: 22) of the coding region of the plasmid pDG8. pDG8 is a

47

high-copy LEU2 yeast vector (YEplac181) containing Gal4 DNA binding domain, histone H4 (amino acids 1-29), Gcn5 F221A mutant allele (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 20 shows the nucleotide sequence (SEQ ID NO: 9) and amino acid sequence (SEQ ID NO: 23) of the coding region of the plasmid pDG28. pDG28 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 21 shows the nucleotide sequence (SEQ ID NO: 10) and amino acid sequence (SEQ ID NO: 24) of the coding region of the plasmid pDG29. pDG29 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, Gcn5 F221A mutant allele (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 22 shows the nucleotide sequence (SEQ ID NO: 11) and amino acid sequence (SEQ ID NO: 25) of the coding region of the plasmid pDG30. pDG30, based on pDG1, the H3 sequence is replaced with a multicloning sequence. The MSC allows insertion of known and putative substrates for Gcn5 in the tethered catalysis/yeast two-hybrid assays.

Figure 23 shows the nucleotide sequence (SEQ ID NO: 12) and amino acid sequence (SEQ ID NO: 26) of the coding region of the plasmid pDG31. pDG31, based on pDG2, the H3 sequence is replaced with a multicloning sequence.

Figure 24 shows the nucleotide sequence (SEQ ID NO: 13) and amino acid sequence (SEQ ID NO: 27) of the coding region of the plasmid pMK485. pMK485 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, tumor suppressor protein p53 amino acids 300-393, Gcn5 (amino acids 18-252), and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 25 shows the nucleotide sequence (SEQ ID NO: 14) and amino acid sequence (SEQ ID NO: 28) of the coding region of the plasmid pMK486. pMK486 is a low-copy TRP1 yeast vector (pODB2) containing Gal4 DNA binding domain, tumor suppressor protein p53 amino acids 300-393, Gcn5 (amino acids 18-252) with the

F221A mutation, and a trimeric HA epitope under the control of yeast ADH1 promoter and ADH1 terminator.

Figure 26 shows the phosphorylation of the carboxyl terminal domain (CTD) by the tethered Kin28 kinase. (A) is a diagrammatic representation of the procedure. (B) is a Westernblot (*see*, Example 11).

Figure 27 shows identification of proteins that interact specifically with the phosphorylated CTD.

Figure 28 shows the autophosphorylation of the histone H3 at the Ser10 residue by the tethered Ipl1 protein kinase.

Figure 29 shows the PIASxα and PIASxβ proteins interact with p53 in an acetylation-dependent and -independent manner.

Figure 30 shows the nucleotide sequence (SEQ ID NO: 29) and amino acid sequence (SEQ ID NO: 30) of the coding region of the plasmid pDG64.

Figure 31 shows the nucleotide sequence (SEQ ID NO: 31) and amino acid sequence (SEQ ID NO: 32) of the coding region of the plasmid pDG65.

Figure 32 shows the nucleotide sequence (SEQ ID NO: 33) and amino acid sequence (SEQ ID NO: 34) of the coding region of the plasmid pMK498.

Figure 33 shows the nucleotide sequence (SEQ ID NO: 35) and amino acid sequence (SEQ ID NO: 36) of the coding region of the plasmid pMK500.

Figure 34 shows the nucleotide sequence (SEQ ID NO: 37) and amino acid sequence (SEQ ID NO: 38) of the coding region of the plasmid pDG502.

DESCRIPTION OF THE INVENTION AND PREFERRED EMBODIMENTS

Although the present invention is not limited by specific theories or mechanisms of action, it is believed that histones are the substrates of multiple PTMs that are critical for probably all DNA-templated processes. Additionally, for example, histone acetyltransfereases (HATs) are targets for oncogenesis. For example, two of the best known mammalian HATs, p300 and CBP, are targets of several oncoproteins, and translocation mutations of p300/CBP and other HATs have been found in certain cancers (Timmermann, S., *et al.*, "Histone acetylation and disease" *Cell Mol Life Sci*

58:728-36, 2001). One normal function of p300/CBP is tied to a machinery controlling DNA damage repair (Rapic-Otrin, V., *et al.*, "Sequential binding of UV DNA damage binding factor and degradation of the p48 subunit as early events after UV irradiation" *Nucleic Acids Res* 30:2588-98, 2002; Tini, M., *et al.*, "Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription" *Mol Cell* 9:265-77, 2002). Oncoproteins may also function through the recruitment of HDACs, leading to acute promyelocytic leukemia, lymphoid oncogenic transformation, and acute myeloid leukemia (Marks, P., *et al.*, "Histone deacetylases and cancer: causes and therapies" *Nat Rev Cancer* 1:194-202, 2001). Therefore, mutations that influence the balance of histone acetylation may have significant roles in carcinogenesis. Currently, more than a dozen synthetic or natural HDAC inhibitors already show tumor inhibition activity in animal models, and at least six of them are being tested in Phase I and II clinical trials (Marks, P., *et al.*, "Histone deacetylases and cancer: causes and therapies" *Nat Rev Cancer* 1:194-202, 2001). Identification and studies of AcBPs in yeast will undoubtedly help identify human oncoproteins and/or tumor suppressors displaying similar affinity. Therefore, one embodiment of the present invention contemplates novel cancer treatments and screening methods based on the compounds and methods of the present invention.

Chromatin Structure

Eukaryotic chromatin provides a structural basis for genomic DNA organization that is essential for packaging the entire genome into the nucleus and chromosome segregation during mitosis and meiosis. In contrast, all DNA-templated processes require appropriate access and even progressing through selective loci by large multi-subunit machineries under specified conditions. How chromatin structure is regulated to meet these two antagonistic needs is a critical question under very active investigation. Several conserved mechanisms control the dynamic characteristics of chromatin, including covalent modifications of histones, selective use of histone variants, and ATP hydrolysis-dependent chromatin remodeling activities (Hayes, J. J., and J. C. Hansen, "Nucleosomes and the chromatin fiber" *Curr Opin Genet Dev*

11:124-9, 2001; Wolffe, A, "Chromatin, structure and function" *Academic Press* 1998; Wolffe, A. P., and J. J. Hayes, "Chromatin disruption and modification" *Nucleic Acids Res* 27:711-20, 1999). Other mechanisms such as DNA methylation and special RNA molecules (*e.g.*, the XIST and small interfering RNAs (siRNAs)), frequently impose more widespread and stable effects on chromatin (Hall, I. M., *et al.*, "Establishment and maintenance of a heterochromatin domain" *Science* 297:2232-2237, 2002; Kelley, R. L., and M. I. Kuroda, "Noncoding RNA genes in dosage compensation and imprinting" *Cell* 103:9-12, 2000; Mlynarczyk, S. K., and B. Panning, "X inactivation: Tsix and Xist as yin and yang" *Curr Biol* 10:R899-903, 2000; Panning, B., and R. Jaenisch, "RNA and the epigenetic regulation of X chromosome inactivation" *Cell* 93:305-308, 1998; Reinhart, B. J., and D. P. Bartel, "Small RNAs correspond to centromere heterochromatic repeats" *Science* 297:1831, 1998; Volpe, T. A., *et al.*, "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi" *Science* 297:1833-1837, 2002). Research in our laboratory revolves around the functional studies of covalent modifications of histones. It has been a challenge to establish the molecular mechanisms by which different histone modifications, or a single modification recurring at different residues of histones, may bring about different biological functions, such as transcriptional regulation of selective genes, DNA replication and chromatin assembly, recombination, and DNA damage repair. The "histone code" hypothesis (Strahl, B. D., and C. D. Allis, "The language of covalent histone modifications" *Nature* 403:41-45, 2000) suggests that histone modifications function as transducing signals to recruit certain proteins to the underlying loci for specific molecular functions. In this proposal, we use histone acetylation as the model to test this hypothesis. Specifically, we have developed a genetic system permitting non-biased screening for protein-protein interactions induced by histone modifications. Functional characterization of the acetylated histone binding proteins (AcBPs) is expected to shed light on the spectrum and the mechanisms of histone acetylation functions. Furthermore, based on the AcBP studies, our research has been extend to other modifications to obtain a panorama view of how chromatin dynamics may be determined by covalent modifications of histones.

51

Core histone N-terminal tails are covalently modified

Core histones H2A, H2B, H3, and H4 are highly conserved proteins essential for chromatin organization. Two molecules of each of the core histones are wrapped around by about 150 basepairs of DNA to form a nucleosome (van Holde, K. E, "Chromatin" *Springer-Verlag* 1989; Wolffe, A., "Chromatin, structure and function" *Academic Press* 1998). Formation of nucleosomes requires extensive histone-histone and histone-DNA interactions occurring mainly within the central histone-fold domain of each core histone (Arents, G., R. W. Burlingame, *et al.*, "The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix" *Proc Natl Acad Sci U S A* 88:10148-10152, 1991; Luger, K., A., *et al.* "Crystal structure of the nucleosome core particle at 2.8 Å resolution" *Nature* 389:251-260, 1997). Each core histone contains an amino-terminal tail and sometimes a carboxyl-terminal extension. Much of the histone tails protrude from the nucleosomal core particle. Crystal structures of nucleosomal core particles indicate that histone tails are not structured and some part of them may interact with adjacent nucleosomes (Hansen, J. C., C. Tse, and A. P. Wolffe, "Structure and function of the core histone N-termini: more than meets the eye" *Biochemistry* 37:17637-17641, 1998; Luger, K., *et al.*, "Crystal structure of the nucleosome core particle at 2.8 Å resolution" *Nature* 389:251-260, 1997; White, C. L., R. K. Suto, and K. Luger, "Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions" *Embo J* 20:5207-5218, 2001). Deleting the amino tail domains of histones H3 and H4, or of the H2A and H2B together causes yeast cell death (Ling, X., *et al.*, "Yeast histone H3 and H4 amino termini are important for nucleosome assembly in vivo and in vitro: redundant and position-independent functions in assembly but not in gene regulation" *Genes Dev* 10:686-699, 1996). These results demonstrate the importance of histone tails in cell viability. However, the direct cause of the cell death remains unclear.

Histone tails are the targets for multiple post-translational modifications including acetylation, methylation, phosphorylation, ubiquitylation and several other less studied chemical changes (Spencer, V. A., and J. R. Davie, "Role of covalent

modifications of histones in regulating gene expression" *Gene* 240:1-12, 1999; van
Holde, K. E, "Chromatin" *Springer-Verlag* 1989; Wolffe, A "Chromatin, structure and
function" *Academic Press*, 1998; Wolffe, A. P., and J. J. Hayes, "Chromatin disruption
and modification" *Nucleic Acids Res* 27:711-720, 1999). Individual and combined

5　　　actions of these covalent modifications may contribute significantly to the general
functions of histone tails. Many of these modifications change the ionic charge of the
highly basic histones. At the first approximation, alteration of the ionic state of
histones can have substantial effects on the compact structure of chromatin which
generally restricts the binding and progression of protein factors. For example, the

10　　Km of the interaction between a highly basic histone H4 tail peptide and
double-stranded DNA is about 10-12 M, whereas acetylation of this peptide decreases
the affinity by a factor of 10^6 (Hong, L., *et al.*, "Studies of the DNA binding
properties of histone H4 amino terminus. Thermal denaturation studies reveal that
acetylation markedly reduces the binding constant of the H4 "tail" to DNA" *J Biol*

15　　*Chem* 268:305-314, 1993). A weakened DNA-histone interaction may allow for better
access of DNA binding and processing factors to find their cognate DNA elements
(Tse, C., T. Sera, A. P. Wolffe, and J. C. Hansen, "Disruption of higher-order folding
by core histone acetylation dramatically enhances transcription of nucleosomal arrays
by RNA polymerase III" *Mol Cell Biol* 18:4629-4638, 1998; Vettese-Dadey, M., *et al.*,

20　　"Acetylation of histone H4 plays a primary role in enhancing transcription factor
binding to nucleosomal DNA *in vitro*" *Embo J* 15:2508-2518, 1996; Workman, J. L.,
and R. E. Kingston, "Alteration of nucleosome structure as a mechanism of
transcriptional regulation" *Annu Rev Biochem* 67:545-579, 1998). On the other hand,
the diversity of histone tail modifications also suggests that multiple mechanisms may

25　　be used by these covalent modifications in different chromatin-related functions
(detailed below). In fact, probably all DNA-templated processes related to chromatin
metabolism are influenced or accompanied by one or more of these modifications.

　　　Many histone modifying enzymes have been identified. Histone
acetyltransferases (HATs) and deacetylases (HDACs) are two families of opposing

30　　enzymes that acetylate and deacetylate histones, respectively (Kuo, M. H., and C. D.

Allis, "Roles of histone acetyltransferases and deacetylases in gene regulation" *Bioessays* 20:615-626, 1998; Peterson, C. L., "HDAC's at work: everyone doing their part" *Mol Cell* 9:921-922, 2002, Roth, S. Y., J. M. Denu, and C. D. Allis, "Histone acetyltransferases" *Annu Rev Biochem* 70:81-120, 2001). Methylation of histones occurs at lysine and arginine residues (van Holde, K. E., "Chromatin" *Springer-Verlag* 1989). Arginine and lysine methyltransferases have been found in different organisms (Jenuwein, T., "Re-SET-ting heterochromatin by histone methyltransferases" *Trends Cell Biol* 11:266-273, 2001). Thus far, no known enzymes actively remove the methyl moiety from a methylated histone (Bannister, A. J., R. Schneider, and T. Kouzarides, "Histone methylation: dynamic or static?" *Cell* 109:801-806, 2002). Several kinases possess the histone phosphorylation activity (De Souza, C. P., *et al.*, "Mitotic histone H3 phosphorylation by the NIMA kinase in *Aspergillus nidulans*" *Cell* 102:293-302, 2000; Hsu, J. Y., *et al.*, "Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes" *Cell* 102:279-291, 2000; Lo, W. S., *et al.*, "Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription" *Science* 293:1142-1146, 2001). Less is certain with regards to the histone phosphatases. Lastly, one histone H2B ubiquitin ligase has been found in the budding yeast (Robzyk, K., J. Recht, and M. A. Osley, "Rad6-dependent ubiquitination of histone H2B in yeast" *Science* 287:501-504, 2000); ubiquitylation of histones is a widely conserved events with functions outside protein degradation (Jason, L. J., *et al.*, "Histone ubiquitination: a tagging tail unfolds?" *Bioessays* 24:166-174, 2002).

Considering the number of modifications each histone tail may have and the size of each histone tails (from around 20 to 60 amino acids), these modifications occur at a fairly high density. Also, lysine residues can be acetylated, methylated, and ubiquitylated. It is thus not surprising that these modifications may influence each other. For example, acetylation at lysine 14 of H3 can be facilitated by phosphorylation at serine 10 (Cheung, P., *et al.*, "Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation" *Mol Cell* 5:905-915, 2000; Lo, W. S., *et al.*, "Phosphorylation of serine 10 in histone

H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at lysine 14" *Mol Cell* 5:917-926, 2002), probably because of an increased affinity between the HATs and the phosphorylated H3 (Lo, W. S., *et al.*, "Phosphorylation of serine 10 in histone H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at

5 lysine 14" *Mol Cell* 5:917-926, 2002). Lysine methylation is found enriched in hyperacetylated loci and is tied to transcriptional activation (Strahl, B. D., *et al.*, "Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1" *Curr Biol* 11:996-1000, 2001). The deacetylase complex, NuRD, is excluded from nucleosomes containing H3 methylated at lysine 4

10 (Zegerman, P., *et al.*, "Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex" *J Biol Chem* 277:11621-11624, 2002), partly explaining how histone acetylation and methylation may be enriched at the same region. Moreover, H3 K4 methylation is completely abolished in a yeast strain where the H2B ubiquitylation is prevented (Dover, J., *et al.*,

15 "Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6" *J Biol Chem* 277:28368-28371, 2002; Sun, Z. W., and C. D. Allis, "Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast" *Nature* 418:104-108, 2002), strongly suggesting that different histone modifications may cross-talk and coordinate each other's action.

20 **Histone acetylation is necessary for DNA-templated nuclear activities**

Acetylation is one of the best studied histone modifications. Genetic and biochemical studies on HATs, HDACs, and on the acetylatable lysine residues have established the roles of histone acetylation in transcriptional regulation (Elgin, S. C. R., and J. L. Workman (ed.), "Chromatin structure and gene expression" *Oxford*

25 *University Press*, 2000: Turner, B. M., "Chromatin and gene regulation" *Blackwell Science*, 2001). Hyper- and hypo-acetylated histones are generally associated with transcriptional activation and repression, respectively. In many cases, transcriptional activators first bind to their target DNA sequences and then recruit coactivators to the promoter (Agalioti, T., *et al.*, "Ordered recruitment of chromatin modifying and

general transcription factors to the IFN-beta promoter" *Cell* 103:667-678, 2000;

Cosma, M. P., T. Tanaka, and K. Nasmyth, "Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter" *Cell* 97:299-311, 1999; Krebs, J. E., *et al.* "Cell cycle-regulated histone acetylation

5   required for expression of the yeast HO gene" *Genes Dev* 13:1412-1421, 1999; Kuo, M. H., *et al.*, "Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription" *Mol Cell* 6:1309-1320, 2000). Many of the transcriptional coactivators possess histone acetylation or chromatin remodeling activities. The HAT then acetylates nucleosomes at the promoter and activates

10   transcription (Kuo, M. H., *et al.*, "Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription" *Mol Cell* 6:1309-1320, 2000; Kuo, M. H.,*et al.*, "Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo" *Genes Dev* 12:627-639, 1998; Parekh, B. S., and T. Maniatis, "Virus infection leads to localized hyperacetylation of histones H3 and H4 at

15   the IFN-beta promoter" *Mol Cell* 3:125-129, 1999). Transcriptional repressors and co-repressors, frequently containing the HDAC activity, function in a similar fashion (Narlikar, G. J., H. Y. Fan, and R. E. Kingston, "Cooperation between complexes that regulate chromatin structure and transcription" *Cell* 108:475-487, 2002). In some cases, acetylation may help other DNA binding factors bind their cognate elements

20   (Krebs, J. E., *et al.*, "Cell cycle-regulated histone acetylation required for expression of the yeast HO gene" *Genes Dev* 13:1412-1421, 1999; Vettese-Dadey, M., *et al.*, "Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro" *Embo J* 15:2508-2518, 1996), or help chromatin remodeling complexes perform their functions (Barbaric, S., *et al.*,

25   "Increasing the rate of chromatin remodeling and gene activation--a novel role for the histone acetyltransferase Gcn5" *Embo J* 20:4944-4951, 2001). However, in most other cases, it is unknown which step(s) of transcriptional activation is directly affected by histone acetylation. Furthermore, deviation from this "acetylation = activation" dogma does exist. For example, mutations of certain HATs actually perturb transcriptional

30   silencing (Sun, Z. W., and M. Hampsey, "A general requirement for the Sin3-Rpd3

histone deacetylase complex in regulating silencing in Saccharomyces cerevisiae" *Genetics* 152:921-932, 1999), suggesting that the transcriptional readout may not be the result of a simple acetyllysine counting mechanism.

Other nuclear activities are linked to acetylation as well. For example, the yeast and human Elongator complexes that are important for transcriptional elongation contain HAT components; the HAT activity is an integral and essential part of the complexes (Kim, J. H., W. S. Lane, and D. Reinberg., "Human Elongator facilitates RNA polymerase II transcription through chromatin" *Proc Natl Acad Sci U S A* 99:1241-1246, 2002; Wittschieben, B. O., *et al.,* "Overlapping roles for the histone acetyltransferase activities of SAGA and elongator *in vivo*" *Embo J* 19:3060-3068, 2000). The yeast NuA3 HAT complex (Sas3 is the catalytic subunit) interacts with Spt16 that is a component of yeast CP (Cdc68/Pob3) (Brewster, N. K., G. C. Johnston, and R. A. Singer, "A bipartite yeast SSRP1 analog comprised of Pob3 and Nhp6 proteins modulates transcription" *Mol Cell Biol* 21:3491-3502, 2001; Evans, D. R., et al., "The yeast protein complex containing cdc68 and pob3 mediates core-promoter repression through the cdc68 N-terminal domain" *Genetics* 150:1393-1405, 1998) and mammalian FACT (Facilitates chromatin transcription) complexes (John, S., *et al.,* "The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)-FACT complex" *Genes Dev* 14:1196-1208, 2000). These two complexes also function in transcriptional elongation. V(D)J joining in immune cells has been suggested to be enhanced by histone hyperacetylation at the recombination signal sequences (McBlane, F., and J. Boyes, "Stimulation of V(D)J recombination by histone acetylation" *Curr Biol* 10:483-486, 2000; McMurry, M. T., and M. S. Krangel, "A role for histone acetylation in the developmental regulation of VDJ recombination" *Science* 287:495-498, 2000), although other data argue for a more important role played by promoter positioning (Sikes, M. L., *et al.,* "Regulation of V(D)J recombination: A dominant role for promoter positioning in gene segment accessibility" *Proc Natl Acad Sci U S A* 99:12309-12314, 2002). For DNA repair, one yeast HAT Gcn5 is important for photoreactivation and nucleotide excision repair of

UV-induced cyclobutane pyrimidine dimers at certain loci (Teng, Y., Y. Yu, and R. Waters, "The Saccharomyces cerevisiae histone acetyltransferase Gcn5 has a role in the photoreactivation and nucleotide excision repair of UV-induced cyclobutane pyrimidine dimers in the MFA2 gene" *J Mol Biol* 316:489-499, 2002). Another yeast HAT, Esa1 complex, is recruited to the double-strand DNA breaks for both nonhomologous end joining repair and a new replication-coupled repair pathway (Bird, A. W., *et al.*, "Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair" *Nature* 419:411-415, 2002). The human p300/CBP acetyltransferase is found associated with the p127 subunit of the UV-damaged DNA binding protein complex (UV-DDB) that is implicated in global genomic nucleotide excision repair (Rapic-Otrin, V., *et al.*, "Sequential binding of UV DNA damage binding factor and degradation of the p48 subunit as early events after UV irradiation" *Nucleic Acids Res* 30:2588-2598, 2002), as well as with the thymine DNA glycosylase that functions in repair of G/T and G/U mismatches (Tini, M., *et al.*, "Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription" *Mol Cell* 9:265-277, 2002).

Multiple HATs, multiple acetylation, and multiple functions

Multiple HATs and HDACs exist in probably all eukaryotes, in sync with the many functions linked to acetylation. These enzymes may display very different substrate specificities (Table 1).

Table 1

Table 1. Predicted protein-protein interactions using the different Autocatalytic baits

	C-A-B-D	C-A-B'-D	C-B-D	C-B'-D	The detected interaction is:
1	+	-	-	-	induced by A modification
2	-	+	-	-	inhibited by A modification

3	+	+	-	-	independent of A modification
4	+	+	+	+	not specific for the protein A

For example, even though Lys14 of H3 seems to be the favored acetylation target for
many HATs, these enzymes may differ from each other in their ability to acetylate
other lysines of H3 or other core histones (Sterner, D. E., and S. L. Berger,
"Acetylation of histones and transcription-related factors" *Microbiol Mol Biol Rev*
64:435-459, 2000). The mammalian p300/CBP HATs acetylate multiple lysines of all
four core histones (Ogryzko, V. V., *et al.*, "The transcriptional coactivators p300 and
CBP are histone acetyltransferases" *Cell* 87:953-959, 1996). Esa1 is an essential HAT
which prefers H2A and H4 (Smith, E. R., *et al.*, "ESA1 is a histone acetyltransferase
that is essential for growth in yeast" *Proc Natl Acad Sci U S A* 95:3561-3565, 1998).
The catalytic subunit of the Elongator complex, Elp3, is able to acetylate all four core
histones in an in-gel activity assay (Wittschieben, B. O., *et al.*, "A novel histone
acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme"
Mol Cell 4:123-128, 1999), whereas the isolated complex acetylates.K14 of H3 and K8
of H4 (Winkler, G. S., *et al.*, "Elongator is a histone H3 and H4 acetyltransferase
important for normal histone acetylation levels *in vivo*" *Proc Natl Acad Sci U S A*
99:3517-3522, 2002). The significance of maintaining different acetylation patterns is
exemplified by several reports that the global acetylation of H3 and H4 establishes a
transcriptionally poised state (Hebbes, T. R., *et al.*, "Core histone hyperacetylation co-
maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal
domain" *Embo J* 13:1823-1830, 1994; Kuo, M. H., *et al.*, "Gcn4 activator targets
Gcn5 histone acetyltransferase to specific promoters independently of transcription"
Mol Cell 6:1309-1320, 2000; Schubeler, D., *et al.*, "Nuclear localization and histone
acetylation: a pathway for chromatin opening and transcriptional activation of the
human beta-globin locus" *Genes Dev* 14:940-950, 2000; Vogelauer, M., J. Wu, N.
Suka, and M. Grunstein, "Global histone acetylation and deacetylation in yeast" *Nature*
408:495-498, 2000), whereas the promoter-specific hyperacetylation of H3 seems to be

59

a more direct cause of transcriptional activation (Kuo, M. H., *et al.*, "Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription" *Mol Cell* 6:1309-1320, 2000; Parekh, B. S., and T. Maniatis, "Virus infection leads to localized hyperacetylation of histones H3 and H4 at the IFN-beta promoter" *Mol Cell* 3:125-129, 1999). Similarly, different yeast HDACs not only show distinct preference on the target acetyllysines, but also differ in the genomic loci to which they are recruited and function (Kurdistani, S. K., *et al.*, "Genome-wide binding map of the histone deacetylase Rpd3 in yeast" *Nat Genet* 31:248-254, 2002; Peterson, C. L., "HDAC's at work: everyone doing their part" *Mol Cell* 9:921-922, 2002; Robyr, D., *et al.*, "Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases" *Cell* 109:437-446, 2002).

Compared with transcriptional regulation, less is known how other nuclear activities may be affected by different acetylation patterns. For example, Esa1 and Gcn5 are important for DNA damage repair via separate pathways (Bird, A. W., *et al.*, "Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair" *Nature* 419:411-415, 2002; Teng, Y., Y. Yu, and R. Waters,."The Saccharomyces cerevisiae histone acetyltransferase Gcn5 has a role in the photoreactivation and nucleotide excision repair of UV-induced cyclobutane pyrimidine dimers in the MFA2 gene" *J Mol Biol* 316:489-499, 2002). Is this functional differentiation a result of the very different histone acetylation patterns generated by these two enzymes? When Gcn5 participates in UV damage repair, does it create an acetylation product identical to that generated during transcriptional activation? If so, how does one acetylation pattern specify different biological functions? If not, do these differences of the acetylation pattern have physiological significance? Furthermore, arginine mutations introduced at selective lysine residues of histone tails appear to cause dissimilar outcomes in gene activity and chromatin assembly (Braunstein, M., *et al.*, "Efficient transcriptional silencing in Saccharomyces cerevisiae requires a heterochromatin histone acetylation pattern" *Mol Cell Biol* 16:4349-5436, 1996; Ma, X. J., *et al.*, "Deposition-related sites K5/K12 in histone H4 are not required for nucleosome deposition in yeast" *Proc Natl Acad Sci U S A* 95:6693-6698, 1998; Mann, R. K., and

60

M. Grunstein, "Histone H3 N-terminal mutations allow hyperactivation of the yeast GAL1 gene *in vivo*" *Embo J* 11:3297-306., 1992; Zhang, W., *et al.*, "Essential and redundant functions of histone acetylation revealed by mutation of target lysines and loss of the Gcn5p acetyltransferase" *Embo J* 17:3155-3167, 1998), indicating that each

5   lysine residue, and likely the acetylation at these sites, may play different roles.

In short, how acetylation of histones controls selective, locus-specific functions remains a mystery. At a broader scale, much less is known as to how histone modifications exert their molecular functions. The present invention is ideal to find answers to this conundrum. We are particularly interested in using the present

10   invention to test whether a histone/nucleosome bearing a particular acetylation pattern may perform specific nuclear functions, and if so, whether these functions are carried out by proteins with specific affinity toward this acetylation pattern. The present invention, a novel "autocatalysis/yeast two-hybrid" method, can identify proteins based on their ability to bind an acetylated histone *in vivo*. These proteins are known to

15   perform distinct biological roles, and hence are likely to link histone acetylation to different chromatin functions. It is contemplated that the present invention can analyze these acetylated histone binding proteins (AcBPs) to see how one particular acetylated histone species may contribute to particular nuclear functions. Furthermore, the present invention can be used to identify other AcBPs that prefer different acetylated

20   histone populations. Functional studies of these AcBPs will likely shed light on the wide spectrum of functions linked to histone acetylation and, importantly, how these functions are performed at a molecular level.

Post-translational Modifications

Proteins are the most versatile macromolecules in living systems and serve

25   crucial functions in essentially all biological processes. Many proteins also function at the intersection between discrete cellular pathways, such as the communication between chromatin integrity surveillance (aka, checkpoint), cell cycle control, and programmed cell death pathways. The appropriate execution within a function and the coordination between different pathways require numerous interactions between

proteins. Stable or transient interactions with selective protein partners are essential for the functions of most, if not all, proteins. Understanding protein-protein interaction at a proteomic scale is now an achievable goal which will ultimately reveal how normal cells function and how malignancies, for example, arise from misregulation of certain cellular activities.

Currently available data suggest that there are at least 10,000 protein-protein interactions among the 6,200 yeast proteins (Uetz, P., Curr Opin Chem Biol 6:57-62, 2002). Such estimation is mainly based on known protein-protein interactions carried out between the "native" or "unmodified" proteins. The total number of protein-protein interactions obviously increases when the genome size increases. That is, human proteins (30,000 - 60,000 are encoded by the human genome) perform a much greater number and combination of distinct protein-protein interactions. On the other hand, numerous proteins contained post-translational modifications (PTMs) in which selective chemical moieties are added to specific amino acid residues of the target proteins after these proteins are synthesized (see below). Evidence showed that PTMs may trigger or prevent protein-protein interactions. Few currently available methods are suitable for detecting such interactions at a global scale. Thus, our current knowledge on proteomic interactions is far from complete unless those interactions requiring specific PTMs are identified and investigated.

Chemical moieties that constitute post-translational modifications include, for example, the acetyl group (acetylation), the methyl group (methylation), the hydroxyl group (hydroxylation), simple and complex sugars (glycosylation), lipids (myristoylation, palmitoylation, etc), phosphate (phosphorylation), ubiquitin (ubiquitylation), etc. The biological significance of PTMs can be exemplified by phosphorylation: There are 120 kinases in yeast (total protein-encoding genes are around 6,000), and ~550 kinases in human (30,000 to 60,000 total genes). It is estimated that 30% of cellular proteins contain covalently bound phosphate (Cohen, P. 2000. Trends Biochem. Sci. 25:596-602). Given the wide variety of PTMs, it is quite possible that most, if not all, aspects of cellular functions require appropriate regulation of PTMs of specific proteins. The molecular consequences of PTMs vary

significantly, including protein stability, intra- and extra-cellular localization, co-factor binding/removal, activation/inactivation of the enzymatic activities of the modified proteins, association/dissociation with other protein factors, and so on. Of these known functions, the potential of a given PTM to recruit or repel a specific protein

5   partner(s) is one of the very important, yet least characterized. Local conformational changes resulted from the PTM, or the chemical moiety itself along with the nearby sequences may solicit protein-protein interactions that are specific for the modified state. Alternatively, an existing protein-protein interaction may be abolished by the PTM. An increasing amount of evidence, mostly obtained by sporadic analyses,

10   supports this notion. However, the lack of an efficient, non-biased genetic method that allows genome-wide identification of such interactions thwarts our full exploration of this territory. In the following sections, a brief review of literature on the involvement of PTMs in protein-protein interactions is given, followed by a detailed description of the Autocatalysis/Two-Hybrid system that provides a versatile and novel solution to

15   this problem.

Phosphorylation of a Tumor Suppressor Protein P53 Recruits Acetyltransferases

Mutations of a tumor suppressor protein p53 have been found in greater than 50% of cancer patients. When cells are exposed to UV, ionizing radiation, and other DNA damaging agents, p53 accumulates in the nucleus and regulates the expression of

20   many genes to arrest the cell cycle so that DNA damages can be repaired. If the damage is too extensive to be repaired, p53 instead triggers apoptosis (programmed cell death) to wipe out the damaged cells so that the mutation will not be passed to progeny cells. A cascade of molecular events, including phosphorylation and acetylation, leads to the accumulation and activation of p53 in the nucleus. Though

25   carried out by distinct enzymes, p53 phosphorylation and acetylation are intimately related to each other in that the phosphorylated p53 binds an acetyltransferase p300/CBP better than without the phosphorylation. p300/CBP then acetylates the carboxyl domain of p53. Meanwhile, p300/CBP recruits yet another acetyltransferase PCAF, which also acetylates p53 within its carboxyl domain. The heavily acetylated

p53 further recruits transcriptional coactivators and activates certain genes for cell cycle arrest (Barlev, *et al.*, *Mol. Cell* 8:1243-1252, 2001).

Histone Acetylation and Methylation Recruit Regulators for Transcriptional Control

5   Histones are the basic protein constituents for eukaryotic genome organization, *i.e.*, the chromatin. Histones serve two opposing functions for chromatin structures. On the one hand, histones condense the chromatin which nucleates the formation of mitotic chromosomes for equal distribution of the two sets of genome to the daughter cells during cell division. The rigid structure of chromosomes renders most genomic

10   loci refractory for nuclear activities such as gene activation. On the other hand, histones undergo a variety of PTMs which control the biochemical and biophysical characteristics of histones and hence the dynamics of chromatin. Many of the histone modifications antagonize the condensing roles of histones so that selective loci are poised for gene activation, recombination, and other nuclear functions. Histone

15   modifications include acetylation, methylation, phosphorylation, ubiquitylation, and some other less studied PTMs. Probably all nuclear DNA-templated processes (i.e. transcriptional regulation, DNA replication, chromatin assembly during cell division, DNA damage repair, and recombination) are affected by one or more histone modifications. The mechanisms by which histone PTMs regulate the underlying locus

20   activity remain large unclear. The "histone code" hypothesis suggests that each specifically modified histone acts a transducing signal to recruit other proteins for different molecular functions. Indeed, the acetylated histones are bound by several transcriptional activators containing the bromodomain (see Example 2), whereas methylated histones are bound by several chromodomain-containing transcriptional

25   repressors. The known functions of histone acetylation and methylation in gene regulation correlate well with the corresponding binding proteins. Furthermore, the Silent Information Regulator protein Sir3 represses transcription by binding the unacetylated histones; acetylation of histones inhibits the binding of Sir3 protein and causes transcriptional de-silencing.

Autocatalysis/Two-Hybrid System to Identify Protein-Protein Interactions Involving PTMs

The above examples clearly indicate that protein-protein interactions may be induced or inhibited by specific post-translational modifications. An efficient and non-biased method that allows for the identification of such interactions will be of immense importance for constructing the proteomic interactions database in any organism. The Autocatalysis/Two-Hybrid system (AC/2H) provides such a method.

The essence of any genetic method deriving from the Yeast Two-Hybrid system to identify protein-protein interactions involving PTMs is the effective creation of a specifically and constitutively modified bait. The current invention is novel in that it generates a specialized bait which has the unique ability to catalyze the desired covalent modification within itself at a specific amino acid residue(s). The presence of the covalent modification within the bait allows protein-protein interactions that is induced by this modification to be identified. Moreover, a counterscreen using an otherwise identical bait but lacks the specific PTM will sort out interactions that are independent of, or are inhibited by the covalent modification under investigation.

The detailed design of the AC/2H system is illustrated in Figure 1. In nature, the protein of interest, A, can be modified by an enzyme, B, in a traditional trans-reaction. The rate of these two proteins to encounter and associate with each other in the environment dictates the efficiency of the catalysis. In contrast, in the AC/2H system, the enzyme and the substrate are encoded as a single protein (or fused to other modules for the purpose of two-hybrid screening, see below), therefore enzyme B catalyzes the modification of A while these two proteins are covalently linked to each other. In other words, for every molecule of the enzyme synthesized, there is a molecule of the substrate within its vicinity. The rate of catalysis can thus reach maximum (the maximal rate of an enzymatic action, Vmax, is defined as that when all enzyme molecules associate with the substrates). As a control, the substrate protein A is also fused to a mutant enzyme B which contains a pre-determined mutation that abolishes the catalytic power of the enzyme (the resulting mutant is denoted B'). The substrate A within the A-B' fusion thus remains unmodified within

the substrate A. When A-B and A-B' are used in two parallel two-hybrid tests (after fusing these two to modules C and D, see legends of Figure 1), proteins that require a modified A for the interaction will display positive reporter readout with the A-B but not the A-B' fusion. On the other hand, if a protein only interacts with the

5   unmodified A, a positive interaction will then be detected by A-B' but not A-B fusion. Proteins that interact with A independently of the latter's modification status will be scored positive in both A-B and A-B'.

Exemplary Uses and of the AC/2H System

As summarized below and without limiting the present invention to any particular use,

10   the power and some of the novel uses of the AC/2H system are several fold. These descriptions below are provided as exemplary only and do not limit the invention in any way.

In one embodiment, it is contemplated that the substrate and the enzyme used in the autocatalysis context can be any known reaction partners. As shown in the

15   examples in the EXPERIMENTAL section, histones H3 and H4 were fused to a histone acetyltransferase Gcn5 which leads to auto-acetylation of both histones by the linked Gcn5. In one embodiment, a substitution of the histone acetyltransferase Gcn5 with other histone modifying enzymes such as Snf1, a known histone H3 kinase, will create an phosphorylated AC bait. Likewise, substituting the histone with other

20   proteins, such as the tumor suppressor protein p53 that is known to be acetylated, phosphorylated, and ubiquitylated, one can identify proteins that interact only with the modified p53 protein. By selecting for those interactions that occur only when p53 is fused to an enzymatically inactivated, but not the wildtype enzyme, p53 interaction partners that are excluded by specific modification(s) of p53 can also be identified. It

25   is thus contemplated that the present invention can detect protein-protein interactions induced or inhibited by a variety of post-translational modifications.

The autocatalytic capacity of the enzyme-substrate fusion of the present invention is not affected by which two-hybrid system is chosen (see Figures 5 and 6). Presently, several methods complement the original Yeast Two-Hybrid system (U.S.

Patent Nos. 5,283,273, 5,468,614 and 5,667,973) in different ways. Some of these derivatives include the Split-Ubiquitin system (Johnsson, *et al.*, "Split ubiquitin as a sensor of protein interactions in vivo" *Proc. Natl. Acad. Sci. USA* 91:10340-10344, 1994; U.S. Patent Nos. 5,503,977 and 5,585,245), Bacterial Two-Hybrid and Multi-Hybrid systems (U.S. Patent No. 6,333,154), and Sos and Ras Recruitment system (Aronheim, 2001, Methods Enzymol 332:260-70). Certain kinds of protein-proteins interactions are not likely to be detected by the Y2H (such as those occurring on or within the membrane), but can be detected by one or more of these derivatives. It is contemplated that the autocatalysis concept can be used in conjunction with these methods and hence maximize our ability to screen for PTM-triggered or perturbed protein-protein interactions (see Example 1). All U.S. Patents referred to in this document are incorporated herein by reference.

In one embodiment, it is contemplated that a substrate can be fused to more than one enzyme such that multiple post-translational modifications can be added to the substrate simultaneously. If certain protein-protein interactions require concomitant PTMs of one of the two interacting proteins, a tandem autocatalytic bait (i.e. a fusion composed of substrate A-enzyme B1-enzyme B2) can be created.

In one embodiment, it is also contemplated that two proteins may interact with each other only when both of them contain specific modifications. One can thus create autocatalytic bait and prey (*i.e.*, substrate A1-enzyme B1 and substrate A2-enzyme B2) and fuse these hybrids to the appropriate two-hybrid modules to test the interaction.

In another embodiment, a Modified RNA binding protein screening using the Autocatalysis concept is contemplated. The ability of certain proteins to interact with selective RNA molecules plays critical roles in a variety of biological functions, such as pre-mRNA splicing, telomerase activity, RNA transport, etc. The yeast Three-Hybrid System (SenGupta, D.J. *et al.*, "A three-hybrid system to detect RNA-protein interactions *in vivo*" *Proc Natl Acad Sci U S A* 93:8496-501, 1996) is a derivative of the Y2H method to detect protein-RNA interactions. Many RNA molecules are known to be modified after synthesis (*i.e.*, post-transcriptional modifications). If the responsible enzyme is known, the Autocatalysis concept can be applied to the

Three-Hybrid system to screen for the proteins that interact with only the appropriately modified RNA molecules (Figure 2). For example, in the Three-Hybrid system, the bait is composed of two hybrid proteins: the first hybrid protein consists of a DNA binding motif and a known RNA-binding protein. The second hybrid, an RNA hybrid, is a fusion of two RNAs: one being the ligand for the RNA-binding protein within the first hybrid, and the second being the RNA of interest to which the interacting proteins are to be screened/tested. The third hybrid is the traditional activation domain fusion. To incorporate the Autocatalysis design to the Three-Hybrid method so that proteins that interact only with the modified RNA can be detected, the RNA modifying enzyme can be fused to the first (protein) hybrid. When the second (RNA) hybrid is recruited to the promoter via interaction with the RNA-binding protein within the first triple-hybrid protein, the fused RNA modifying enzyme can modify the bait RNA. If the third hybrid contains the cognate binding protein, positive interactions can then be detected. A parallel fusion with the mutant RNA modifying enzyme will yield a negative result on the interaction test.

In one embodiment, it is contemplated that the DNA is known to be modified under certain conditions. For example, methylation of DNA is the basis for the prokaryotic restriction system. In eukaryotes, DNA methylation has been linked to gene regulation and developmental control. Methylated DNA recruits selective proteins that repress transcription. As shown in Figure 3, the target DNA sequence is engineered to the proximity of the target sequence to which the bait protein binds (*e.g.*, the UASgal that binds the GDBD). The DNA modifying enzyme is fused to the DNA binding module so that it will be brought to the target DNA element via the protein-DNA interaction between UASgal and GDBD. If the AD fusion contains a protein that binds the modified DNA element, transcriptional activation of the reporter gene will be detected. On the other hand, if the mutant DNA modifying enzyme is used, or if the modification target DNA in omitted, no interaction will be detected.

i) Autophosphorylation of CTD by Kin28 and Ctk1 Kinases.

Phosphorylation is the best known PTM. Protein-protein interactions triggered

68

by phosphorylation of one of the two interacting partners have been reported in different systems, and it is very likely that many more such interactions exist in divergent cellular functions. It is well known that 14-3-3 proteins bind phosphoserine/phosphothreonine proteins, whereas SH2 and PTB proteins bind

5   phosphotyrosine in a context-dependent manner. On the other hand, proteins lacking the 14-3-3, SH2, or PTB modules may be novel proteins that bind specifically phosphorylated protein targets. One example is the WW domain protein Ess1 that interacts with the phosphorylated Carboxyl Terminal Domain (CTD) of the largest subunit of the RNA polymerase II in eukaryotes (Myers, *et al.*, "Phosphorylation of

10   RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1" Biochemistry 40:8479-86, 2001). At least two other proteins also interact with the phosphorylated CTD (Ho, et al, "The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II" J Biol Chem 273:9577-85, 1998;

15   McCracken, et al, "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II" Genes Dev 11:3306-18, 1997). CTD phosphorylation is intimately associated with transcriptional elongation (Riedl, T., and J. M. Egly, "Phosphorylation in transcription: the CTD and more" Gene Expr 9:3-13, 2000). Several autoimmune diseases are associated with

20   auto-antibodies against the CTD (Dahmus, M. E., "Phosphorylation of the C-terminal domain of RNA polymerase II" Biochim Biophys Acta 1261:171-82, 1995). Therefore, the phosphorylated CTD is an excellent model for search for proteins that bind phosphorylated proteins, with the known phosphoprotein-binding modules or not.

The CTD of the largest subunit of RNA polymerase II is composed of tandem

25   repeats of a heptapeptide Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Different kinases phosphorylate different residues. For example, Ser2 is phosphorylated by Csk1, and Ser5 is the preferred target for Kin28 and several other protein kinases (Bensaude, et al, "Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD)" Biochem Cell Biol 77:249-55, 1999; Keogh, et al, "Kin28 is found within TFIIH and a

30   Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop

phosphorylation" Mol Cell Biol 22:1288-97, 2002; Murray, et al, "Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase" Mol Cell Biol 21:4089-96, 2001). In one embodiment, it is contemplated that CTD, phosphorylated at Ser2 or Ser5, acts as the target for protein-protein interactions. It is also contemplated that CTD-Csk1 and CTD-Kin28 are created and ligated in-frame to GDBD-HA construct to create autophosphorylation baits. Thus, the methodologies described in Examples 1-5 are employed to characterize the phosphorylation status of the fused CTD within the autocatalysis context. Additionally, it is contemplates that genetic screens are used to identify proteins that function as phosphorylated CTD-binding proteins.

ii) Automethylation of Histones H3 and H4.

Transcriptional activation is associated with histones H3 and H4 Arg3 methylation as well as H3 Lys4 methylation (Strahl, et al, "Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1" Curr Biol 11:996-1000, 2001; Wang, et al, "Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor" Science 293:853-7, 2001), whereas transcriptional repression and silencing are associated with histone H3 Lsy9 methylation (Lachner, et al, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins" Nature 410:116-20, 2001; Nakayama, et al, "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly" Science 292:110-3, 2001). Lys9 methylation is known to recruit chromodomain-containing proteins (Lachner, et al, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins" Nature 410:116-20, 2001). Although histones methylated at arginine residues have not been shown to bind other proteins, arginine methylation in SmD1 and SmD3 was shown to be recognized by the Survivor of Motor Neurons (SMN) protein (Friesen, et al, "SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets" Mol Cell 7:1111-7, 2001). The search for extra proteins that interact specifically with methylated proteins is thus of high significance in both basic and clinical research. In

the yeast Saccharomyces cerevisiae, at least three lysine methyltransferases modify histone H3: Set1 (Lys4) (Briggs, et al, "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae" Genes Dev 15:3286-95, 2001; Bryk, et al, " Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism" Curr Biol 12:165-70, 2002), Set2 (Lys36) (Strahl, et al, "Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression" Mol Cell Biol 22:1298-306, 2002), and Dot1 (Lys79) (Dlakic, M., "Chromatin silencing protein and pachytene checkpoint regulator Dot1p has a methyltransferase fold" Trends Biochem Sci 26:405-7, 2001; van Leeuwen, *et al.*, "Dot1p modulates silencing in yeast by methylation of the nucleosome core" Cell 109:745-56, 2001). In addition, Arg3 of H4 is methylated by Rmt1 (Lacoste, *et al.*, "Disruptor of Telomeric Silencing-1 Is a Chromatin-specific Histone H3 Methyltransferase" J Biol Chem 277:30421-4, 2002). In one embodiment of the present invention, it is contemplated that the collection of different methylated histone species provides an excellent model to screen for methylated histone binding proteins. Toward this end, we have constructed H3-Set1, H3-Set2, H3-Dot1, and H4-Rmt1 fusion constructs. In short, these fusion fragments can be inserted in-frame with GDBD-HA and expressed in yeast for immunochemical characterization for the desired modifications. Enzymatically inactive versions of each enzyme can be included in the counter-screening constructs as the negative control for subsequent genetic screening. When the automethylation is confirmed, genetic screening is carried out.

Advantages of the AC/2H System

The current invention offers several advantages over existing methods. In one embodiment, the enzyme catalyzes the substrate modification in cis (*i.e.*, autocatalytically). Although the present invention is not limited to any particular theory, it is believed that the enzyme acts at its maximal rate and efficiency. This is clearly different and much more preferable than the typical, trans-reactions of most natural or artificial protein modifications. In another embodiment, the autocatalytic

71

enzyme-substrate fusion can be expressed in its natural host where the opposing enzymes (*e.g.*, deacetylases vs. acetyltransferases, phosphatases vs. kinases, etc.) are present. While pleiotropic effects are frequently seen when protein modifying enzymes are overexpressed (especially when the opposing enzyme is absent), the AC/2H system

5   does not need to over- or ectopically express the enzyme. It is thus much less likely that adverse effects may result from the Autocatalysis setting. In yet another embodiment, the inclusion of the catalytically inactivated enzyme in a parallel chimeric protein fusion provides an ideal control for protein-protein interactions that do not require the modification of the bait protein. In yet another embodiment, the reversal in

10   the genetic screening criteria can reveal protein interactions that are perturbed by a the bait modification. In yet another embodiment, the use of tandem array of different protein modifying enzymes in the autocatalytic baits may provide baits possessing multiple modifications. One single construct is thus sufficient for the bait creation and the target protein screening. In yet another embodiment, the bait bearing the specific

15   chemical modification can be a protein (Figure 1), an RNA (Figure 2), and a DNA (Figure 3).

EXPERIMENTAL

The following examples serve to demonstrate certain aspects of the present invention and do not limit it in any way.

20   **Example 1**

Evidence of Autocatalysis in Two Different Two-Hybrid Systems and Two Different Organisms. The concept and feasibility of Autocatalysis was tested using an array of fusion proteins composed of histones H3, H4 (as the substrates) and the prototypic histone acetyltransferase, Gcn5. The detailed overview of this system is

25   shown in Figure 4.

To first test if the autocatalysis actually is feasible, histone H3 (amino acids 2 to 60) was fused to Gcn5 (the catalytic domain, amino acids 1 to 252). In the first setting, H3 was fused to the glutathione S-transferase (GST), followed by the Gcn5

catalytic domain, and by the Ras protein. The Ras protein is a part of the bait used in the Ras Recruitment System, an alternative to Stanley Fields' Yeast Two-Hybrid system (Aronheim, A., *Methods Enzymol* 332:260-70, 2001) (U.S. Patent No. 5,776,689). The GST is an epitope tag allowing efficient purification of the fusion protein. In addition, a point mutation of Gcn5, F221A (Kuo, M.-H., *et al.*, "Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo" Genes Dev 12:627-639, 1998) was also used to create parallel, catalytically inactive fusion proteins. The F221A significantly diminishes the enzymatic activity of Gcn5, and hence provides an un-acetylated histone H3 bait for counter-screening. In addition, the H3-GST-GCN5-Ras DNA construct was inserted in a bacterial expression vector and a yeast vector. Autocatalysis can thus be tested in proteins synthesized in either E. coli or yeast. The fusion proteins were synthesized and purified from E. coli or yeast, resolved by SDS-PAGE, and analyzed by western blots using two antibodies. The first antibody recognizes preferentially the acetylated histone H3, whereas the second antibody recognizes the unacetylated histones. As shown in Figure 5A, the bacterial H3-Gcn5 fusion clearly demonstrates autoacetylation whereas the mutant Gcn5 function is not detectably acetylated under the same condition. Figure 5B shows very similar results from fusion proteins derived from yeast. These results clearly indicate that the autocatalysis is not dependent on the host cells.

Furthermore, a fusion between Gcn5 and yet another histone, H4, was also created. The acetylation status of H4-GST-Gcn5-Ras, expressed and purified from yeast, is shown in Figure 5C. Again, the western results clearly indicate the acetylation of H4 by the wildtype Gcn5 but not the F221A mutant.

Figure 6 shows the autocatalysis when H3-Gcn5 was expressed within the prototypic Yeast Two-Hybrid system context. In this setting, H3-Gcn5 was inserted between the Gal4 DNA binding domain (GDBD) and the hemaglutinin (HA) epitope tag (Figure 12, pDG1 [SEQ ID NOS: 1 and 15]; Figure 13, pDG2 [SEQ ID NOS: 2 and 16]; Figure 16, pDG5 [SEQ ID NOS: 5 and 19] and; Figure 17, pDG6 [SEQ ID NOS: 6 and 20]) and expressed in yeast. Yeast proteins were prepared and immunoprecipitated by an antibody against the HA tag. The immunoprecipitated

materials were thus subjected to western analyses using the acetylated H3-specific antibody mentioned above. The western blot results are very clear: when H3 was fused to the wildtype Gcn5, it was acetylated efficiently; however, when H3 was fused to the F221A mutant Gcn5 fragment, no acetylation was detected.

In short, autoacetylation is clearly achieved in two different two-hybrid systems and in two different organisms. In sharp contrast, the mutant Gcn5 fusion fails to catalyze the autoacetylation. Therefore, the concept of autocatalysis and the use of a catalytically inactive mutation to create the unmodified bait for counter-screening has been proved feasible.

Example 2

Confirmation of An Interaction between Acetylated Histones and the Bromodomain of the PCAF Protein. This Example is to show that AC/Y2H recapitulates bromodomain-acetylated histone interaction in vivo.

To further confirm that the AC/2H can identify protein-protein interactions that require specific PTMs, the GDBD-H3-Gcn5-HA constructs were used in the Yeast Two-Hybrid genetic tests. In the Y2H system, the expression of one of the three reporter genes reveals the protein-protein interactions (James, *et al., Genetics* 144:1425-1436, 1996). The first reporter is the bacterial lacZ gene under the control of GAL7 promoter. Positive interactions are indicated by elevated ß-galactosidase activity. The second reporter is the HIS3 gene under the control of the GAL1 promoter. When *HIS3* gene is upregulated by positive protein-protein interactions, yeast cells display significant resistance to the chemical 3-amino-1,2,4-triazole (3-AT) and survive in the absence of histidine (His). The third reporter construct is the ADE2 gene fused to the promoter of GAL2 gene. Yeast cells gain the ability to survive in the absence of adenine (Ade) when protein-protein interactions exist between the bait and the prey proteins.

In the first test of AC/Y2H, it was asked whether a previously reported interaction between acetylated histones and the PCAF bromodomain can be detected by our system. This interaction was identified by biochemical means (Dhalluin, C., *et al.,*

74

"Structure and ligand of a histone acetyltransferase bromodomain" *Nature* 399:491-496, 1999; Jacobson, R. H., *et al.*, "Structure and function of a human TAFII250 double bromodomain module" *Science* 288:1422-1425, 2000). To test this *in vivo* with the AC/Y2H system, the bromodomain of the PCAF protein was fused to the Activation Domain (AD) of the Gal4 transcriptional activator. The AD-PCAF expression construct was transformed into different yeast strains bearing a variety of AC baits. The two-hybrid interaction was assessed by measuring the b-galactosidase activity. As shown in Figure 7, the PCAF bromodomain interacts with H4-Gcn5 (wildtype, pDG3) fusion but not the mutant Gcn5 counterpart (pDG4). A weaker interaction was detected between the PCAF bromodomain and the H3-Gcn5 fusion (pDG1). Again, the mutant Gcn5 fusion (pDG2) showed a negligible level of lacZ expression. The activation of lacZ caused by GDBD-H3/H4-Gcn5(wt) alone (bars 1 and 3) is an anticipated background level of transcription. This is because tethering Gcn5, a transcriptional coactivator, to the promoter had been shown to induce modest transcription (Marcus, G. A., *et al.*, "Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors" Embo J 13:4807-4815, 1994). In conclusion, the AC/2H system is able to detect a previously reported protein-protein interaction that requires a specific post-translational modification. Additionally, these data are the first *in vivo* evidence that the highly conserved bromodomain is indeed able to interact with specific acetylated histones.

Example 3

Identification of Novel Acetylated Histone Binding Proteins Using the AC/Y2H Methodology. This Example shows that AC/Y2H using a modified chromatin component (*e.g.*, acetylated histone H3) identifies three chromatin-related proteins Cac1, Rmt1, and Rpm2.

To definitively test whether the AC/2H system is suitable for genetic screening, the GDBD-H3-Gcn5-HA (pDG1) was tested in two different formats of Y2H. The first approach uses a high-throughput screening method (Uetz, P., *et al.*, "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae" *Nature* 403:623-7,

2000). In this approach, protein-protein interactions were tested in about 6,000 yeast

strains simultaneously. Each of these 6,000 strains expresses a unique chimeric protein

that contains the Gal4 activation domain (AD) and one of the about 6,000 open reading

frames (ORFs). The high-throughput Y2H approach uses a robot to cross each of these

5   AD-fusion yeast haploid strains to the one that contained either the

GDBD-H3-Gcn5(wildtype)-HA (*i.e.*, pDG1) or the GDBD-H3-Gcn5(F221A)-HA (*i.e.*,

pDG2) expression plasmid. The ability of each of the diploid strain after the crossing

(now both the bait and one of the 6,000 prey proteins are present in the same diploid

cell) to grow in a medium lacking histidine (-His) or adenine (-Ade) was assessed.

10   Candidates that showed positive interactions with the wildtype Gcn5 fusion but not the

mutant Gcn5 hybrid were sorted out and tested again for the growth in different media.

As seen in Figure 8A, Rmt1, Cac1 and Exo84 allowed yeast cells to grow in -His

medium in the presence of the acetylated H3 bait (resulting from the H3-wildtype Gcn5

fusion), but not of the unacetylated H3 (*i.e.*, the H3-Gcn5 F221A fusion). When these

15   same cells were tested under a more stringent condition (-Ade medium), only Rmt1-AD

fusion allows H3-wildtype Gcn5 bait-bearing cells to grow. These results indicate that

Cac1, Exo84, and Rmt1 possess intrinsic affinity to acetylated histone H3, and that the

Rmt1 may interact with the acetylated H3 at the highest affinity among the three.

In addition, we also tested the interaction between Rpd3-AD fusion and

20   H3-Gcn5 chimeric proteins. Rpd3 is a histone deacetylase (Rundlett, S. E., *et al.*, *Proc*

Natl Acad Sci U S A 93:14503-14508, 1996; Taunton, *et al.*, *Science* 272:408-411,

1996). HDACs are the most obvious AcBPs. The enlarged photo insert in Figure 8

shows that a weak interaction can be detected between the acetylated H3 and Rpd3.

This weak association is likely resulting from the constant turnover and transient nature

25   of enzyme-substrate interactions.

In the second test, a yeast genomic DNA library with AD fusion was used to

screen for acetylated histone H3 binding proteins. This is the "traditional" type of Y2H

screen that is being used in numerous labs nowadays. In this test, yeast cells were

sequentially transformed with the GDBD-H3-Gcn5(wildtype)-HA construct and the

30   AD-yeast DNA library. Yeast transformants were tested for their ability to grow in the

absence of adenine for strong interactions solicited by the H3 acetylation. From this screen, another protein, Rpm2, was found to be a strong acetylated H3 interacting protein (Figure 9).

Previously reported results indicate that the identification of Cac1, Rmt1, and Rpm2 as the novel acetylated histone binding protein is very significant:

CAC1/RLF2: Cac1 (Kaufman, *et al., Genes Dev* 11:345-357, 1997), or Rlf2 (Enomoto, *et al., Genes Dev* 11:358-370, 1997), is the largest subunit of the yeast chromatin assembly factor complex-I (CAF-I). The activity of CAF-I is conserved from yeast through human (Kaufman, *et al., Cell* 81:1105-1114, 1995; Kaufman, *et al., Genes Dev* 11:345-357, 1997). It is thought that CAF-I binds and delivers newly synthesized histones H3 and H4 to DNA replication forks for nucleosome assembly. Curiously, the human CAF-I interacts with H3/H4 in a tail independent manner Kaufman, *et al., Cell* 81:1105-1114, 1995; Verreault, *et al., Cell* 87:95-104, 1996). My AC/Y2H result thus shows an unsuspected, acetylation-dependent role played by CAF-I. Indeed, CAF-I also participates in (Enomoto, *et al., Genes Dev* 12:219-232, 1998; Monson *et al., Proc Natl Acad Sci U S A* 94:13081-13086, 1997). The silencing functions are likely mediated through the Sas2-containing HAT complex, SAS-I (Meijsing, *et al., Genes Dev* 15:3169-3182, 2001). Further, H4 K16 mutation confers the same de-silencing phenotype caused by the sas2 null mutation (Meijsing, *et al., Genes Dev* 15:3169-3182, 2001), linking histone acetylation to Cac1 functions. Furthermore, deleting CAC1 causes defects in repairing UV-damaged DNA (Game, *et al., Genetics* 151:485-497, 1999). Recent data also showed that CAF-I and Hir proteins associate with the kinetochore and are important for centromere functions (Sharp, *et al., Genes Dev* 16:85-100, 2002).

RMT1/HMT1: Rmt1 (protein arginine methyltransferase), or Hmt1 (hnRNP methyltransferase 1) (Henry, *et al., Mol Cell Biol* 16:3668-3678, 1996), transfers the methyl moiety from S-adenosyl methionine to specific arginine residues of certain proteins. Known substrates for Rmt1 include Npl3 (Henry, *et al., Mol Cell Biol* 16:3668-3678, 1996) and Nab2 (Green, *et al., J Biol Chem* 277:7752-7760, 2002). Nab2 and Nlp3 function in splicing and mRNA transport. The Nab2 function depends

on its methylation by Rmt1 (Green, *et al.*, *J Biol Chem* 277:7752-7760, 2002). Arginine methylation of several transcriptional activators is important for gene activation (Mowen, *et al.*, *Cell* 104:731-741, 2001; Zhu, *et al.*, *J Biol Chem* 277:35787-35790, 2002). Rmt1 also methylates H4 at Arg3 in vitro (Lacoste, *et al.*, *J Biol Chem* 277:30421-30424, 2002). In mammals, H4 Arg3 methylation is important for transcriptional induction by steroid hormones (Bauer, *et al.*, *EMBO Rep* 3:39-44, 2002; Ma, *et al.*, *Curr Biol* 11:1981-1985, 2001; McBride, *Cell* 106:5-8, 2001; Wang, *et al.*, *Science* 293:853-857, 2001), although a similar phenomenon has not been seen in yeast. Further, deleting RMT1 does not appreciably diminish H4 Arg3 methylation (Lacoste, *et al.*, *J Biol Chem* 277:30421-30424, 2002), indicating that Rmt1 may perform functions other than methylating the bulk histone H4.

Rpm2: Rpm2 (Ribonuclease P in mitochondria) was first identified as the protein subunit of the mitochondrial RNase P (Morales, *et al.*, *Proc Natl Acad Sci U S A* 89:9875-9879, 1992; Dang and Martin, *J Biol Chem* 268:19791-19796, 1993). Recent proteomic data suggest that Rpm2 may form a complex with a nuclear protein Hrr25 (Gavin, *et al.*, *Nature* 415:141-147, 2002). Hrr25 is a protein kinase that perform a variety of nuclear functions, including DNA damage repair (DeMaggio, *et al.*, *Proc Natl Acad Sci U S A* 89:7008-7012, 1992; Ho, *et al.*, *Proc Natl Acad Sci U S A* 94:581-586, 1997; Hoekstra, *et al.*, *Science* 253:1031-1034, 1991). Affinity purification of Hrr25 also identified histone H4 as another interacting (Gavin, *et al.*, *Nature* 415:141-147, 2002). It is thus possible that Rpm2 interacts with acetylated histones and brings Hrr25 to the target loci for specific functions, such as repair the damages of the underlying loci.

In conclusion, results obtained from the high-throughput and the AD library screens indicate clearly that the Autocatalysis/Two-Hybrid system creates specifically and constitutively modified protein baits in vivo that are suitable for genetic test of protein-protein interactions involving specific post-translational modifications. The AC/2H design thus provides significant improvement over the existing genetic methods for protein-protein interactions.

Example 4

This Example is to show activation domain library screening with H3-Gcn5 AC baits. Although the high-throughput Y2H method of the prior art has uncovered many insightful protein-protein interactions, there seems to be a high rate of false negatives in this approach (Auerbach, D., *et al.*, "The post-genomic era of interactive proteomics: Facts and perspectives" *Proteomics* 2:611-623; Uetz, P., "Two-hybrid arrays" *Curr Opin Chem Biol* 6:57-62, 2002). For example, two independent, yet methodologically very similar Y2H genome-wide screens showed surprisingly small overlap (Ito, T., *et al.*, "A comprehensive two-hybrid analysis to explore the yeast protein interactome" *Proc Natl Acad Sci U S A* 98:4569-4574, 2001; Uetz, P., *et al.*, "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae" *Nature* 403:623-627, 2000). Many previously documented interactions were not picked up by screen. Several explanations are considered. First, it is common that protein-protein interactions are more easily detectable when small domains are used, probably due to the removal of potential interference from the rest of the protein. The current (high-throughput) method uses the entire ORFs for AD fusion. Second, expressing the entire ORF of certain genes may cause adverse effects on growth, hence making the detection of interactions involving these ORFs less likely.

To compensate for the limitation of the high-throughput method, we initiated a traditional Y2H library screen. A library of yeast genomic DNA fragments fused to the Gal4 activation domain (AD) was acquired (James, P., J. Halladay, and E. A. Craig, "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast" *Genetics* 144:1425-1436, 1996) and transformed into the PJ69-4a yeast strain harboring either the GDBD-H3-Gcn5(wt)-HA or GDBD-H3-Gcn5(F221A)-HA construct. The PJ69-4a strain is the same one used in the high-throughput screen. In this strain, three reporter genes are under the control of UASgal: HIS3 and ADE2 respectively confer histidine and adenine prototrophy, and lacZ allows colorimetric quantitation of the transcription and, accordingly, the relative strength of the interaction. It has been shown that ADE2 is a much more stringent reporter than HIS3 and generates significantly less false positives (James, P., J. Halladay, and E. A. Craig,

79

"Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast" *Genetics* 144:1425-1436, 1996). We thus use adenine prototrophy as the primary criterion to screen for AcBPs.

Thus far, ~ 30,000 AD fusion transformants (~ 3 x coverage of the yeast genome) were obtained and replica plated to adenine omission plates. 162 (with wt Gcn5-H3 fusion) and 25 (with F221A Gcn5-H3 fusion) clones were confirmed to be ADE$^+$. Six of these AD plasmids were propagated and purified from *E. coli*, and shown by restriction mapping to contain distinct yeast DNA inserts. As depicted in Figure 4A, these clones were then re-transformed into the parental strain bearing one of several Gcn5 fusion derivatives. These control plasmids can quickly weed out undesired interacting partners. Figure 4B shows that two candidates (clones 5 and 1) interact exclusively with H3-Gcn5(wt), whereas clone 6 confers a ADE$^+$ phenotype whenever the bait contains the wildtype Gcn5; the remainder three did not repeat the ADE$^+$ phenotype in any combination (not shown) and will not be studied further. The identity of these two AD fusion, and further tests of the rest of the putative AcBPs are being pursued at the time of submitting this proposal.

Example 5

This Example shows autoacetylation of a tumor suppressor protein, p53, by physically linked acetyltransferases. Additionally, this example shows that the acetyltransferase Gcn5 is able to mediate autocatalysis when a non-histone protein (p53) is included in the autocatalysis construct. It also indicates that other acetyltransferases, such as p300, may acetylate the fused p53 at different lysine residues. The tumor suppressor protein p53 plays a critical role in determining cell fate in response to DNA damage, nucleotide depletion, hypoxia, and several other genotoxic stresses. These stresses trigger a series of changes in p53 leading to the stabilization and activation of p53 in the nucleus. Activated p53 induces or inhibits the expression of more than 150 genes, many of which are essential for growth, cell cycle control, and apoptosis. The ultimate function of p53 is to commit cells to either DNA damage repair or apoptosis such that mutations are prevented from being passed on to progeny cells. It is

estimated that 50 % of all human cancers are linked to loss-of-function mutations in p53 that result in uncontrolled cellular proliferation. Moreover, suppression of p53 activity in tumor cells can cause tumor relapse after chemotherapy. Interestingly, "superactive" p53 mutants that are predicted to provide enhanced genomic surveillance

5   can cause premature cellular senescence. Therefore, p53 maintains a delicate balance between normal cell proliferation and aging.

A landmark event associated with p53 activation is the post-translational modification (PTM) of p53, most notably acetylation (Gu, W., and R. G. Roeder, "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal

10   domain" *Cell* 90:595-606, 1997) and phosphorylation (Wang, Y., and C. Prives, "Increased and altered DNA binding of human p53 by S and G2/M but not G1 cyclin-dependent kinases" *Nature* 376:88-91, 1995). Like numerous other covalently modified proteins, the exact molecular functions played by various p53 modifications are largely unknown. One likely possibility is that these modifications effect

15   downstream events such as transcriptional activation or repression of p53 target genes. We hypothesize that critical protein-protein interactions that mediate p53 function are controlled by site-specific modifications. For example, interactions between p53 and a downstream protein factor, such as a transcriptional co-activator, may be mediated by a modification at a specific p53 site. Alternatively, some p53 modifications may serve as

20   "repellants" to displace factors that normally associate with p53 in its unmodified, inactive state. Thus, it is possible that site-specific modifications are responsible for determining cellular fate, specifically, to proliferate, apoptose, or senesce. p53-GCN5 fusion: PCAF acetylates K320 of p53 upon UV treatment. The catalytic domains of PCAF and yeast Gcn5 share 56 % identity and 71% similarity. Furthermore, the

25   proline 16 of histone H3 that is critical for Gcn5-H3 interaction is conserved in p53 (p53: QPK320KKPLD; H3: TGG14KAPRK) (Rojas, *et al.*, "Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide" *Nature* 401:93-8, 1999). It is thus possible that K320 can be acetylated within the context of p53-yGcn5 chimera as well. To test this possibility, the H3 fragment used in Examples 1-9 have been

30   replaced with amino acids 300 to 393 of p53, and the resultant fusion proteins (with the

wildtype and the mutant Gcn5) were immunoprecipitated from yeast extracts, followed by western analyses to test the acetylation status at Lys320. The results are shown in Figure 10. It is clear that Lys320 of p53, though a non-histone protein, is effectively acetylated by the fused Gcn5 protein, but not the mutant Gcn5. These results firmly establish the feasibility of autoacetylation of general proteins, providing the appropriate enzymes are included in the autocatalytic baits. p53-p300 fusion: p300/CBP acetylates p53 in response to UV and IR treatment. The acetylation sites have been mapped to K372, 373, 381 and 382, with K373 and 382 being the major targets (Gu, W., and R. G. Roeder, "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain" *Cell* 90:595-606, 1999; Liu, *et al.*, "p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage" *Mol Cell Biol* 19:1202-9, 1999). To see whether p300/CBP can be used in p53 autocatalysis, we will follow the strategy stated above to create p53-CBP fusion proteins. In the meantime, a point mutant, F1451A will be included in a parallel construction as the counter selection. F1451 is at the position equivalent to F221 of the yeast Gcn5 and the F1451A mutant loses its ability to acetylate histones and to activate transcription (Martinez-Balbas, *et al.*, "The acetyltransferase activity of CBP stimulates transcription" *EMBO J* 17:2886-93, 1998).

Example 6

This Example shows the identification of human proteins that are able to interact with p53 in acetylation-dependent and -independent manners. Additionally, this example shows that the p53 protein, when fused to the wildtype or mutant Gcn5 acetyltransferase, recruits certain human proteins. As seen in Figure 11, a human cDNA-activation domain (AD) library was transformed into yeast two-hybrid strains containing the GDBD-p53-Gcn5(wt)-HA bait. The transformants were tested for their ability to survive in the absence of adenine. Activation of the ADE2 reporter gene resulted from positive two-hybrid interactions allows cells to form colonies. More then 350,000 yeast transformants were screened and several candidates were obtained. These candidates were further tested for two-hybrid interactions with

GDBD-p53-Gcn5(wt)-HA, GDBD-p53-Gcn5(mutant)-HA, and GDBD-Gcn5(wt)-HA. Three classes of interactions were observed: Class I represents those that only interact with the wildtype Gcn5 fusion of p53 (i.e. the acetylated p53 protein); Class II represents those that interact with both wildtype and mutant Gcn5 fusion of p53 (*i.e.*, p53 interactors independent of the acetylation status); Class III represent those that interact with both wildtype Gcn5-p53 fusion as well as the wildtype Gcn5 alone.

Based on these results, it is highly likely that certain proteins indeed function as acetylated-p53 binding proteins. These Class I proteins may play roles in relaying the p53 functions in transcriptional regulation, cell cycle arrest, and apoptosis. They may also conduct functions in turning over the activated p53 protein when the need of p53 no longer exists.

Example 7

This Example shows the autophosphorylation of the CTD by Kin28 and Ctk1 kinases. Phosphorylation is the best known PTM. Protein-protein interactions triggered by phosphorylation of one of the two interacting partners have been reported in different systems, and it is very likely that many more such interactions exist in divergent cellular functions. It is well known that 14-3-3 proteins bind phosphoserine/phosphothreonine proteins, whereas SH2 and PTB proteins bind phosphotyrosine in a context-dependent manner (see, above). On the other hand, proteins lacking the 14-3-3, SH2, or PTB modules may be novel proteins that bind specifically phosphorylated protein targets. One example is the WW domain protein Ess1 that interacts with the phosphorylated Carboxyl Terminal Domain (CTD) of the largest subunit of the RNA polymerase II in eukaryotes (Myers, *et al.*, "Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1" *Biochemistry* 40:8479-86, 2001). At least two other proteins also interact with the phosphorylated CTD (Ho, *et al.*, "The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II" *J Biol Chem* 273:9577-85, 1998; McCracken, *et al.*, "5'-Capping enzymes are targeted to pre-mRNA by binding to the

phosphorylated carboxy-terminal domain of RNA polymerase II" *Genes Dev*
11:3306-18, 1997). CTD phosphorylation is intimately associated with transcriptional
elongation (Riedl, T., and J. M. Egly, "Phosphorylation in transcription: the CTD and
more" *Gene Expr* 9:3-13, 2000). Several autoimmune diseases result from

5   auto-antibodies against the CTD (Dahmus, M. E., "Phosphorylation of the C-terminal
domain of RNA polymerase II" *Biochim Biophys Acta* 1261:171-82, 1995). Therefore,
the phosphorylated CTD is an excellent model for search for proteins that bind
phosphorylated proteins, with the known phosphoprotein-binding modules or not.

The CTD of the largest subunit of RNA polymerase II is composed of tandem

10   repeats of a heptapeptide Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. Different kinases
phosphorylate different residues. For example, Ser2 is phosphorylated by Csk1, and
Ser5 is the preferred target for Kin28 and several other protein kinases (Bensaude, *et
al.*, "Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD)"
Biochem Cell Biol 77:249-55, 1999; Keogh, *et al.*, "Kin28 is found within TFIIH and a

15   Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop
phosphorylation" *Mol Cell Biol* 22:1288-97, 2002; Murray, *et al.*, "Phosphorylation of
the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase"
Mol Cell Biol 21:4089-96, 2001). To see if CTD, phosphorylated at Ser2 or Ser5, may
act as the target for protein-protein interactions, it is contemplated that CTD-Csk1 and

20   CTD-Kin28 can be created and ligated in-frame to GDBD-HA construct to create
autophosphorylation baits. Methodologies described in Examples 1-5 can thus be
employed to characterize the phosphorylation status of the fused CTD within the
autocatalysis context. Genetic screens will follow to identify proteins that function as
phosphorylated CTD-binding proteins. (Ho, *et al.*, "The guanylyltransferase domain of

25   mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal
domain of RNA polymerase II" *J Biol Chem* 273:9577-85, 1998; McCracken, *et al.*,
"5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated
carboxy-terminal domain of RNA polymerase II" *Genes Dev* 11:3306-18, 1997; Myers,
et al., "Phosphorylation of RNA polymerase II CTD fragments results in tight binding

30   to the WW domain from the yeast prolyl isomerase Ess1" *Biochemistry* 40:8479-86,

2001).

Example 8

This Example shows the automethylation of histones H3 and H4. Additionally, this example indicates the potential use of the AC/2H method in identifying proteins that bind specifically methylated histones. The past two years have seen spectacular explosion of interest in histone methylation and its role in transcriptional regulation. Transcriptional activation is associated with histones H3 and H4 Arg3 methylation as well as H3 Lys4 methylation (Strahl, *et al.*, "Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1" *Curr Biol* 11:996-1000, 2001; Wang, *et al.*, "Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor" *Science* 293:853-7, 2001), whereas transcriptional repression and silencing are associated with histone H3 Lsy9 methylation (Lachner, *et al.*, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins" *Nature* 410:116-20, 2001; Nakayama, *et al.*, "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly" *Science* 292:110-3, 2001). Lys9 methylation is known to recruit chromodomain-containing proteins (Lachner, *et al.*, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins" *Nature* 410:116-20, 2001). Although histones methylated at arginine residues have not been shown to bind other proteins, arginine methylation in SmD1 and SmD3 was shown to be recognized by the Survivor of Motor Neurons (SMN) protein (Friesen, *et al.*, "SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets" *Mol Cell* 7:1111-7, 2001). The search for extra proteins that interact specifically with methylated proteins is thus of high significance in both basic and clinical research. In the yeast Saccharomyces cerevisiae, at least three lysine methyltransferases modify histone H3: Set1 (Lys4) (Briggs, *et al.*, "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae" *Genes Dev* 15:3286-95, 2001; Bryk, *et al.*, " Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S. cerevisiae by a Sir2-independent mechanism" *Curr*

Biol 12:165-70, 2002), Set2 (Lys36) (Strahl, *et al.*, "Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression" *Mol Cell Biol* 22:1298-306, 2002), and Dot1 (Lys79) (Dlakic, M., "Chromatin silencing protein and pachytene checkpoint regulator Dot1p has a methyltransferase fold" *Trends Biochem Sci*

5   26:405-7, 2001; van Leeuwen, *et al.*, "Dot1p modulates silencing in yeast by methylation of the nucleosome core" *Cell* 109:745-56, 2001). In addition, Arg3 of H4 is methylated by Rmt1 (Lacoste, *et al.*, "Disruptor of Telomeric Silencing-1 Is a Chromatin-specific Histone H3 Methyltransferase" *J Biol Chem* 277:30421-4, 2002). The collection of different methylated histone species provides an excellent model to

10   screen for methylated histone binding proteins (MHBPs). Toward this end, we have initiated the construction of H3-Set1, H3-Set2, H3-Dot1, and H4-Rmt1 fusion constructs. In short, these fusion fragments will be inserted in-frame with GDBD-HA and expressed in yeast for immunochemical characterization for the desired modifications. Enzymatically inactive versions of each enzyme will be included in the

15   counter-screening constructs as the negative control for subsequent genetic screening. When the automethylation is confirmed, genetic screening for the MHBPs will be carried out.

TABLE 2

Table 2. Examples of autocatalytic substrate-enzyme fusion

Substrate	Enzyme	PTM	Note
Histone H3	Gcn5	Acetylation	Examples 1-4
Histone H4	Gcn5	Acetylation	Examples 1-4
p53	Gcn5	Acetylation	Example 5
p53	p300/CBP	Acetylation	Example 5
Histone H3	Set1	Methylation (lysine)	Example 7
Histone H4	Rmt1	Methylation (arginine)	Example 7
Histone H3	Set2	Methylation (lysine)	Example 7
Histone H3	Dot1	Methylation (lysine)	Example 7
CTD	Kin28	Phosphorylation	Example 6
CTD	Ctk1	Phosphorylation	Example 6
Histone H3	Snf1	Phosphorylation	
p53	PIAS	Sumoylation	
Histone H2B	Rad6	Ubiquitylation	

5

10

15

Example 9

Autoacetylation of A Tumor Suppressor Protein, p53, by A Physically Linked Acetyltransferase. This example shows that the acetyltransferase Gcn5 is able to mediate autocatalysis when a non-histone protein (p53) is included in the

5   autocatalysis construct. It also indicates that other acetyltransferases, such as p300, may acetylate the fused p53 at different lysine residues.

The tumor suppressor protein p53 is believed to play a critical role in determining cell fate in response to DNA damage, nucleotide depletion, hypoxia, and several other genotoxic stresses. These stresses trigger a series of changes in p53

10   leading to the stabilization and activation of p53 in the nucleus. Activated p53 induces or inhibits the expression of more than 150 genes, many of which are essential for growth, cell cycle control, and apoptosis. The ultimate function of p53 is to commit cells to either DNA damage repair or apoptosis such that mutations are prevented from being passed on to progeny cells. It is estimated that 50 % of all human cancers are

15   linked to loss-of-function mutations in p53 that result in uncontrolled cellular proliferation. Moreover, suppression of p53 activity in tumor cells can cause tumor relapse after chemotherapy. Interestingly, "superactive" p53 mutants that are predicted to provide enhanced genomic surveillance can cause premature cellular senescence. Therefore, p53 maintains a delicate balance between normal cell proliferation and

20   aging.

A landmark event associated with p53 activation is the post-translational modification (PTM) of p53, most notably acetylation (Gu, W., and R. G. Roeder, "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain" Cell 90:595-606, 1997) and phosphorylation (Wang, Y., and C. Prives,

25   "Increased and altered DNA binding of human p53 by S and G2/M but not G1 cyclin-dependent kinases" Nature 376:88-91, 1995). Like numerous other covalently modified proteins, the exact molecular functions played by various p53 modifications are largely unknown. One likely possibility is that these modifications effect downstream events such as transcriptional activation or repression of p53 target genes.

We have found that that critical protein-protein interactions that mediate p53 function are controlled by site-specific modifications. For example, interactions between p53 and a downstream protein factor, such as a transcriptional co-activator, may be mediated by a modification at a specific p53 site. Alternatively, some p53 modifications may serve as "repellants" to displace factors that normally associate with p53 in its unmodified, inactive state. Thus, it is possible that site-specific modifications are responsible for determining cellular fate, that is, to proliferate, apoptose, or senesce.

p53-GCN5 fusion: PCAF acetylates K320 of p53 upon UV treatment. The catalytic domains of PCAF and yeast Gcn5 share 56 % identity and 71 % similarity. Furthermore, the proline 16 of histone H3 that is critical for Gcn5-H3 interaction is conserved in p53 (p53: QPK320KKPLD; H3: TGG14KAPRK) (Rojas, *et al.*, "Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide" Nature 401:93-8, 1999). It is thus possible that K320 can be acetylated within the context of p53-yGcn5 chimera as well. To test this possibility, the H3 fragment used in Examples 1-3 have been replaced with amino acids 300 to 393 of p53, and the resultant fusion proteins (with the wildtype and the mutant Gcn5) were immunoprecipitated from yeast extracts, followed by western analyses to test the acetylation status at Lys320. The results are shown in Figure 10. It is clear that Lys320 of p53, though a non-histone protein, is effectively acetylated by the fused Gcn5 protein (pMK485), but not the mutant Gcn5 (pMK486). These results firmly establish the feasibility of autoacetylation of general proteins, providing the appropriate enzymes are included in the autocatalytic baits.

p53-p300 fusion: p300/CBP acetylates p53 in response to UV and IR treatment. The acetylation sites have been mapped to K372, 373, 381 and 382, with K373 and 382 being the major targets (Gu, W., and R. G. Roeder, "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain" Cell 90:595-606, 1999; Liu, et al, "p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage" Mol Cell Biol 19:1202-9, 1999). To see whether p300/CBP can be used in p53 autocatalysis, we will follow the strategy stated above to create p53-CBP fusion proteins. In the meantime, a point mutant,

F1451A will be included in a parallel construction as the counter selection. F1451 is at the position equivalent to F221 of the yeast Gcn5 and the F1451A mutant loses its ability to acetylate histones and to activate transcription (Martinez-Balbas, et al, "The acetyltransferase activity of CBP stimulates transcription" EMBO J 17:2886-93, 1998).

5

Example 10

Identification of Human Proteins Interacting with p53 in Acetylation-Dependent and -Independent Manners.

This example shows that the p53 protein, when fused to the wildtype or mutant Gcn5 acetyltransferase, recruits certain human 10 proteins.

The p53 protein acetylated at Lys320 by the linked wildtype Gcn5 enzyme (Figure 10, pMK485) was subjected to Y2H screening using a human HeLa cell cDNA library fused to the Gal4 activation domain. Two-hybrid interactions were revealed by the ability of yeast cells to grow in the absence of adenine (-Ade plates). More then 15 350,000 yeast transformants were screened and several candidates were obtained. Figure 11 shows that three classes of interactions were observed: Class I represents those that only interact with the wildtype Gcn5 fusion of p53 (*i.e.*, an acetylated p53); Class II represents those that interact with both wildtype and mutant Gcn5 fusion of p53 (*i.e.*, p53 interactors independent of the acetylation status); Class III represent those 20 that interact with both wildtype Gcn5-p53 fusion as well as the wildtype Gcn5 alone (pDG28).

These results show that the class I proteins function as acetylated-p53 binding proteins. These proteins may play roles in relaying the p53 functions in transcriptional regulation, cell cycle arrest, and apoptosis. They may also conduct functions in turning 25 over the activated p53 protein when the need of p53 no longer exists. The class II proteins represent general p53 interacting proteins. The class III proteins may represent human acetylated histone binding proteins as the wildtype Gcn5 protein tethered to the promoter region may acetylate adjacent histones that recruit the human acetylated histone binding protein-AD fusion to activate the downstream ADE2 gene.

Example 11

Phosphorylation of the carboxyl terminal domain(CTD) by the tethered Kin28 kinase. As shown in Figure 26A, the CTD (consisting of three tandem copies of YSTPSPS) was fused to the Gal4 DNA-binding domain and the wildtype Kin28

5 [SEQ ID NO: 35], or a E54Q catalytically inactive mutant Kin28, and the HA epitope [SEQ IN NO: 37]. As shown in Figure 26B, the fusion proteins were isolated and subjected to western analyses using an antibody specific for CTD phosphorylated at the fifth residue (Ser5). The immunoblot shows that the CTD can be phosphorylated by the wildtype Kin28 protein, whereas the mutant Kin28 fusion is recognized significantly

10 weaker by this antibody, indicating the lack of phosphorylation in this fusion protein.

Example 12

Identification of proteins that interact specifically with the phosphorylated CTD. As shown in Figure 27, the yeast two-hybrid screens were conducted using Kin28 alone [SEQ ID NO: 33], CTD fused to the wildtype Kin28 [SEQ ID NO: 35],

15 and CTD fused to the mutant Kin28 [SEQ ID NO: 37], as the baits. Proteins that interact specifically with the CTD-Kin28 (wildtype) but neither of the other two baits are considered phosphorylated CTD-interacting proteins. Yeast strains containing one of the three baits and a variety of preys (activation domain fusion proteins) were tested for their ability to grow in the presence of different concentration of 3-AT. The ability

20 to grow in such medium indicates stable interactions between the bait(s) and the prey(s). The following proteins (numbered 1-7) are considered phosphorylated CTD-interacting proteins: Fcp1 (a phosphatase known to act on phosphorylated CTD), Ssn8 (or Srb11, a component of the RNA polymerase II holoenzyme), Tfb3 (a component of the RNA polymerase II holoenzyme), Whi2 (a protein involved in

25 cellular growth and a component of a protein phosphatase complex containing the Psr1 catalytic subunit), and three novel proteins (YMR181c, YPL229w, and YDR428c) whose functional links to CTD phosphorylation are novel. In addition, previously known (i.e. Ccl1 and Pcl10) and unknown (YDL100c) Kin28-interacting proteins were also identified in this screen (A-C). Several putative phosphorylated CTD interacting

proteins are not labeled due to the current lack of the sequence information.

Example 13

Autophosphorylation of the histone H3 at the Ser10 residue by the tethered Ipl1 protein kinase. As shown in Figure 28, the histone H3 amino acids 1-59 were fused to the Gal4 DNA binding domain, the wildtype and a catalytically inactive kinase Ipl1, and the HA epitope tag [SEQ ID NO: 29]. The mutations of Ipl1 (E152Q V153L) [SEQ ID NO: 31] completely inactivate the catalytic ability of this enzyme. The fusion proteins were expressed and purified from yeast and subjected to western analyses using an antibody specific for the H3 peptide phosphorylated by the Ser10 position. The western data showed that H3, when fused to the wildtype Ipl1, can be easily recognized by the anti-phosphorylated H3 antibody (anti-H3.Pi). On the other hand, the mutant Ipl1 fusion, though expressed at a significantly higher level than its wildtype counterpart (compare the first and second lanes with anti-HA antibody, left panel), its staining by the phosphorylation-specific antibody is weaker than the wildtype Ipl1 fusion. These results confirm that the H3-Ipl1 (wildtype) autophosphorylates at the H3 Ser10 position.

Example 14

The PIASxα and PIASxβ proteins interact with p53 in an acetylation-dependent and -independent manner. The two proteins identified in the yeast two-hybrid screen shown in Figure 11 were sequenced and found to be PIASxα and PIASxβ (class I and II, respectively). As shown in Figure 29, to demonstrate the physical interaction in a biochemical means, the p53 was expressed as a GST fusion protein and purified from bacteria. A recombinant acetyltransferase, PCAF (the orthologue of the yeast Gcn5 protein), was purified and used to acetylate the p53 protein. The p53 protein, treated with the PCAF for Lys320 acetylation, along with the untreated, unacetylated counterpart, were immobilized to the glutathione beads and incubated with 35S-labelled, in vitro translated PIASxα and PIASxβ proteins. The unbound proteins were washed extensively and the final products bound to the

glutathione beads via interaction with the p53 (acetylated or unacetylated) were analyzed by SDS-PAGE and visualized by fluorography. The results show that the PIASxα interacts preferentially with the acetylated p53, whereas the PIASxβ associates with p53 regardless of its acetylation status. These biochemical results are completely consistent with the yeast two-hybrid growth tests shown in Figure 11, providing the definitive evidence that these two PIAS proteins display distinctive affinity for p53 depending on its acetylation status.

As can be seen from the forging, the present invention provides novel compounds and methods for the detection of interactive proteins wherein such interaction is dependent on one or more post translational modifications.